APPROVAL SHEET

Title of Dissertation: Save the Best for Last: Predictors of Negative Time Preference and Saving the Best for Last

Name of Candidate: Mariana I. Castillo
Doctor of Philosophy, 2020

Dissertation and Abstract Approved: John C. Borrero
Dr. John C. Borrero
Associate Professor
Applied Developmental Psychology

Date Approved: 3/12/20

ABSTRACT

Title of Document: SAVE THE BEST FOR LAST: PREDICTORS OF NEGATIVE TIME PREFERENCE AND SAVING THE BEST FOR LAST

Mariana I. Castillo, Ph.D., 2020

Directed By: Professor John C. Borrero, Ph.D., Department of Psychology

Extensive research suggests that people generally exhibit positive time preference, preferring immediate over delayed outcomes. If positive time preference is normative, a person should typically prefer a worsening sequence (i.e., start with the best outcome, and end with the worst). However, people typically exhibit negative time preference (i.e., preferring an improving series of events) when the choice involves a sequence of outcomes. This dissertation consists of 5 studies exploring features of sequences that may promote negative time preference (i.e., saving the best for last; STBFL). In the first study, undergraduates responded via an online survey. Part 1 was a replication and extension of procedures described by Loewenstein and Prelec (1991). Response patterns like those of Loewenstein and Prelec were observed, in that the percentage of participants who STBFL decreased when the interval between activities in the sequence increased. In Part 2 participants were surveyed about their preference for the order in which they would experience hypothetical outcomes with sequences of different sizes. As array size increased, the percentage of participants who STBFL decreased. Next, three studies were conducted, looking at outcome category as a predictor of negative time preference. First, 192 undergraduates responded to questions involving categorically-different outcomes (e.g., noxious stimuli, food, exercise, schoolwork, leisure). A smaller percentage of participants STBFL relative to prior studies, but the percentage was highest when sequences involved noxious stimuli or food. Second, we examined the correspondence between 8 college students' preference for the order in which they would experience sequences of categorically-different outcomes when those were hypothetical versus real. There was strong correspondence in ranks assigned to hypothetical and real outcomes, but more variability in sequences generated. Third, we assessed preschoolers' preference for sequences. With academic items, 2 of 4 participants chose to STBFL. With leisure items, none STBFL. The final study evaluated whether individuals with intellectual and developmental disabilities (IDD) would STBFL in a Multiple Stimulus Without Replacement (MSWO) preference assessment. With food and toys, 1 of 4 participants STBFL. In addition to replicating a seminal study, this dissertation provides proof of phenomenon in previously unstudied populations, including preschoolers and individuals with IDD.

SAVE THE BEST FOR LAST:
PREDICTORS OF NEGATIVE TIME PREFERENCE
AND SAVING THE BEST FOR LAST

By

Mariana I. Castillo

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore County, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Dedication

To my amazing daughters: Sofia and Cristina. Thank you for teaching me to focus on the beauty of life's little moments even in the face of incredible stress. May you continue being powerful, strong, kind, and beautiful humans, and may you have the strength to always follow your dreams.

Fia y Titi: "las amo hasta el infinito y más allá, ida y vuelta dos veces, hoy…"

Acknowledgements

I would like to express my heartfelt gratitude to Dr. John C. Borrero, my advisor and mentor. Your energy and passion for science inspired me to pursue a Ph.D. Your patience, encouragement and guidance helped me to actually complete it. You taught me the importance of good and effective communication of my science. That it is not just the facts that matter, but the story we tell about them. Thank you for being a great role model, and for teaching me to become a better scientist, behavior analyst, teacher, and person.

My sincere appreciation goes to my committee members: Drs. Griffin Rooker, Susan Sonnenschein, Shuyan Sun, and Jolene Sy. I am grateful for the guidance, support, and feedback you provided since the start of this project. Thank you Dr. Rooker for all your guidance, training, and for letting me borrow the expensive sensory testing equipment! Dr. Sun, thank you for your patience and support with the statistical analyses.

Thank you to the PROMISE team, the Dissertation House, and Dr. Carter-Veale, for helping me stay on track and motivated. Thanks for keeping it real!

Thank you to my colleagues and former lab mates: Amber Mendres-Smith, Michelle Frank-Crawford, and Jessica Becraft. Thank you for paving the road ahead of me and for sharing your lessons learned along the way. I would also like to acknowledge the research assistants who assisted at various stages of this project: Clark Elliot, Yasmeen Alajmi, Sakinah Rushdan, and Jocelyn Ochoa-Garcia.

I would like to acknowledge the study sites. Thank you to the Y-preschool at UMBC for allowing me to conduct part of my research there. Special thanks to The Shafer Center and The Faison center for their invaluable collaboration.

I offer my heartfelt appreciation to my family who provided their love, support and (long-distance) encouragement throughout this journey.

I cannot end this acknowledgement section without expressing my deepest gratitude to my loving, patient, and supportive husband and to my daughters. Conrad, where do I start? Thank you for your expert programming skills. Thank you for your unending support, encouragement, and “tough love” throughout this whole process. Thank you for taking on so many additional responsibilities to give me the time and space I needed to complete this dissertation. Thank you for all the amazing meals, and for helping me to step back, breath, and remember to have a life outside of school and work.

To my daughters, who always inspire me to be better, thank you.

Table of Contents

List of Tables

List of Figures

General Introduction

Planning typically requires that one choose among alternative responses that produce differential outcomes[1]. Even scheduling decisions in the short-term (e.g., food, play, chores, work, daily appointments) involves choosing between alternative sequences because many events that take up time cannot be executed simultaneously, and thus cannot be rescheduled without modifying the order of other activities. Many important choices, whether those involve short or long-term outcomes, require organizing or scheduling a given set of events into a preferred temporal order. For example, one may need to arrange different academic tasks during a school day, schedule exercise sessions during a busy week, decide about the timing of unpleasant medical treatments, or decide the order of events requiring discretionary spending (e.g., the timing of an expensive vacation relative to home improvement projects). How these sequences of outcomes are scheduled or ordered has implications for productivity, health, financial stability, satisfaction, and much more.

Intertemporal choice refers to "decisions in which the moment of choice and the associated consequences are separated in time" (Lucci, 2013, p. 1). Extensive research across species and within humans of different characteristics has found that in intertemporal choice involving single-outcome prospects (i.e., in which only one outcome is selected from the array), delayed outcomes are generally valued less (Ainslie, 1975), and, conversely, sooner outcomes are preferred. In economics, this is referred to as *positive time preference* (Olson & Bailey, 1981).

[1] Experiencing consequences (outcomes) requires some behavior, but for present purposes the emphasis will be placed on the result of those responses.

If positive time preference is evinced on all of one's choices, then when asked to schedule a set of outcomes, one should typically prefer a worsening sequence, in a sense the more "impulsive" choice (i.e., starting with the best outcome and ending with the worst outcome). Numerous studies have independently investigated this matter and have shown that when a choice is viewed as being part of a sequence of outcomes, people typically exhibit *negative time preference* (i.e., they prefer an improving series of events; Loewenstein & Prelec, 1991; Magen, Dweck, & Gross, 2008).

In other words, when the decision frame highlights the individual components of the choice (e.g., would you like to spend time with your friends or work now?), people tend to prefer experiencing the best sooner (i.e., positive time preference; Frederick, Loewenstein, & O'Donoghue, 2002). Yet, when the decision frame emphasizes the sequential aspect of the choice, or a series of unavoidable outcomes (e.g., you can choose the order in which you complete your work assignment and spend time with friends. Which would you like to do first?), people tend to prefer an improving sequence, in a manner "delaying gratification" (i.e., negative time preference; Loewenstein & Prelec, 1991; Magen et al., 2008).

Research has documented a preference for improving sequences across a wide variety of outcomes including money (e.g., Magen et al., 2008), pain (e.g., Ariely & Carmon, 2000), annoying sounds (e.g., Ariely & Zauberman, 2000), restaurant meals (e.g., Loewenstein & Prelec, 1993), health outcomes (e.g., Chapman, 2000), hypothetical scenarios that describe differing affective experiences (e.g., Drolet, Lau-Gesk, & Scott, 2011), and news (e.g., Marshall & Kidd, 1981).

In addition to a variety of outcomes, preference for improving sequences has been shown across a variety of procedural variations (see aforementioned studies); however, this is not true for all studies (e.g., Frederick & Loewenstein, 2008). Targeted research on meaningful parameters (e.g., sequence size, category of outcome) is needed to determine the robustness of preference for improving sequences.

This dissertation consists of a series of studies in which we explored features of sequences that may promote negative time preference (i.e., saving the best for last). A total of five studies separated into three papers, as follows: the first study is described in Paper 1, the next three studies are described in Paper 2 and, the last study is described in Paper 3.

This investigation started with a convenience sample in a human operant laboratory, limited to hypothetical experiences. As such, Paper 1, entitled “Save the Best for Last I: Young Adults Demonstrate Negative Time Preference - A Replication and Extension,” consists of a two-part study investigating whether the timing of choices and the number of outcomes in a sequence, influence participants’ choice to save the best for last. In Part 1 we replicated and extended procedures described by Loewenstein and Prelec (1991, section II) with college students responding to hypothetical questions on an online survey. Next, in Part 2 we surveyed college students about their preference for the order in which they would experience hypothetical outcomes with sequences of different sizes (e.g., 3 activities to sequence or 8).

The next step was to study categorical features of sequences and extend it to an investigation of hypothetical versus real outcomes with two different samples. Thus, Paper 2, entitled “Save the Best for Last II: Whether One Saves the Best for Last

Depends on Outcome Category," describes three studies that sought to expand on existing research that has evaluated categorical features of outcomes in a sequence and their effect on time preference. First, in Study 1, we surveyed college students about their preference for the order in which they would like to experience hypothetical outcomes with sequences of categorically-different outcomes. Specifically, we aimed to study the prevalence of college students' saving the best for last in sequences of categorically-different hypothetical outcomes. For example, does the likelihood of saving the best for last differ when the sequences involve leisure items, school work, food, exercise, noxious experiences, or a mix of all categories? Next, in Study 2 we sought to determine whether there is correspondence between college students' preference for sequence order when outcomes are hypothetical versus real. Finally, in Study 3, we sought to expand the investigation to a younger sample. Specifically, we aimed to determine whether typically-developing preschool children save the best for last when scheduling sequences of school tasks, food, exercise, stories, leisure activities, or a mix of all categories.

Our last study in the series was related to a more applied problem, involving the identification of preferred items to be used as potential reinforcers in behavior interventions to either teach new skills, or eliminate severe problem behavior. Therefore, Paper 3 is entitled "Save the Best for Last III: Some Children Save the Best for Last in the MSWO" and consists of a single study. We believe that if positive time preference exerts control over selection, by extension, one could infer that when presented with an array of items, and prompted to choose one, items that are selected first should be the most preferred. This arrangement is characterized by the multiple-stimulus without replacement (MSWO) preference assessment described by DeLeon and Iwata (1996), a preference assessment procedure used to identify potential reinforcers for behavioral intervention. Preference hierarchies are inferred based on the order of selection, with the item selected first ranked as the highest preferred, the item selected second as the second highest preferred, and so on, with the item selected last as the least preferred. It is possible, however, that the most preferred item may not be correctly identified in an MSWO, if individuals exhibit negative time preference (i.e., if they choose to save the best for last). There is some scattered evidence to suggest that this phenomenon may occur, as has been reported in the literature (Soldberg, Hanley, Layer, & Ingvarsson, 2007), in numerous conference poster presentations (including Becerra & Fahmie, 2014; Litchmore, Ivy, & Weaver, 2014; Pendharkar, Bourret, Nuzzolilli, & Upshaw, 2017; Roath & Fritz, 2015), and in our own clinical practice. The prevalence of the "saving the best for last" phenomenon in the MSWO, however, is unclear and there are currently no published explicit studies of the phenomenon in behavior analytic literature.

In particular, the purpose of our last study, described in Paper 3, was to determine whether individuals with intellectual and developmental disabilities who speak in full sentences will save their most preferred item for last when presented with an array of five items and given a chance to access each in a sequence of their own choosing, in an MSWO arrangement. Following the MSWO, items selected first and last were then assessed under progressive ratio schedules to determine reinforcer efficacy.

Finally, the last chapter provides a general discussion, synthesis, and conclusions. Through these chapters, we will present research on time preference in sequences across the research continuum, with different samples (i.e., college students, typically-developing preschoolers, and children with intellectual and developmental disabilities) and methods. I believe this series of studies will contribute to a thoroughgoing understanding of the roles that specific parameters of sequence choices may have on time preference. As suggested in more detail throughout the following chapters, these findings may serve as a foundation for interpreting and restructuring performance in a variety of settings including organizational, educational, medical, and clinical settings.

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82*, 463-496. doi:10.1037/h0076860

Ariely, D., & Carmon, Z. (2000). Gestalt characteristics of experiences: The defining features of summarized events. *Journal of Behavioral Decision Making, 13*, 191-201. doi: 10.1002/(SICI)1099-0771(200004/06)13:2<191::AID-BDM330>3.0.CO;2-A

Ariely, D., & Zauberman, G. (2000). On the making of an experience: The effects of breaking and combining experiences on their overall evaluation. *Journal of Behavioral Decision Making, 13*, 219-232. doi: 10.1002/(SICI)1099-0771(200004/06)13:2<219::AID-BDM331>3.0.CO;2-P

Becerra, L. & Fahmie, T. A. (2014, May). *A review and analysis of the consistency of MSWO assessments.* Poster presented at the 40th Annual Convention of the Association for Behavior Analysis International, Chicago, IL.

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

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29*, 519-533. doi: 10.1901/jaba.1996.29-519

Drolet, A., Lau-Gesk, L., & Scott, C. (2011). The influence of aging on preferences for sequences of mixed affective events. *Journal of Behavioral Decision Making, 24*, 293-314. doi:10.1002/bdm.695

Frederick, S., & Loewenstein, G. (2008). Conflicting motives in evaluations of sequences. *Journal of Risk and Uncertainty, 37,* 221-235. doi:10.1007/s11166-008-9051-z

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature, 40*, 351-401. doi:10.1257/002205102320161311

Litchmore, M., Ivy, J. W., & Weaver, M. T. (2014, May). *Reinforcer strength: An examination of assumptions of preference assessment results.* Poster presented at the 40th Annual Convention of the Association for Behavior Analysis International, Chicago, IL.

Loewenstein, G., & Prelec, D. (1991). Negative time preference. *The American Economic Review, 81*, 347-352.

Loewenstein, G. F., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review, 100*, 91-108.

Lucci, C. R. (2013). Time, self, and intertemporal choice. *Frontiers in Neuroscience, 7,* 1-5. doi: 10.3389/fnins.2013.00040

Magen, E., Dweck, C. S., & Gross, J. J. (2008). The hidden-zero effect: Representing a single choice as an extended sequence reduces impulsive choice. *Psychological Science, 19,* 648-649. doi:10.1111/j.1467-9280.2008.02137.x

Marshall, L. L., & Kidd, R. F. (1981). Good news or bad news first? *Social Behavior and Personality, 9,* 223-226. doi:10.2224/sbp.1981.9.2.223

Olson, M., & Bailey, M. (1981). Positive time preference. *Journal of Political Economy, 89,* 1-25. Retrieved from http://www.jstor.org.proxy-bc.researchport.umd.edu/stable/1837349

Pendharkar, T., Bourret, J. C., Nuzzolilli, A., & Upshaw, S. (2017, May). *Saving the best for last in MSWO preference assessments.* Poster presented at the 43rd Annual Convention of the Association for Behavior Analysis International, Denver, CO.

Roath, C. & Fritz, J. N. (2015, May). *Order of selections in the Multiple-Stimulus Without Replacement preference assessment.* Symposium presented at the 41st Annual Convention of the Association for Behavior Analysis International, San Antonio, TX.

Soldberg, K. M., Hanley, G. P., Layer, S. A., & Ingvarsson, E. Y. (2007). The effects of reinforcer pairing and fading on preschoolers' snack selections. *Journal of Applied Behavior Analysis,* 40, 633-644. doi: 10.1901/jaba.2007.633–644

Save the Best for Last I: Young Adults Demonstrate Negative Time Preference-

A Replication and Extension

Mariana I. Castillo, Shuyan Sun, Michelle A. Frank-Crawford, and John C. Borrero

UMBC

Completion of this study and preparation of this manuscript were supported in part by grants from the UMBC Graduate Student Association.

The authors would like to thank Dr. Griffin Rooker, Dr. Jolene Sy, and the anonymous reviewers for their feedback on earlier drafts of this manuscript.

Direct correspondence to J. C. Borrero, Department of Psychology, UMBC, 1000 Hilltop Circle, Baltimore, Maryland 21250 (e-mail: jborrero@umbc.edu).

Mariana Castillo ORCID 0000-0002-7433-4324

Shuyan Sun ORCID 0000-0002-7512-2753

Michelle Frank-Crawford ORCID 0000-0003-4319-6317

John Borrero ORCID 0000-0003-0413-0892

Abstract

Generally, immediate outcomes are preferred to delayed outcomes, and in economics, this is referred to as positive time preference. If positive preference is normative, when asked to schedule a set of outcomes, a person should typically prefer a worsening sequence (i.e., choose to start with the best outcome, and end with the worst outcome). Several studies have shown that when a choice is among a *sequence* of outcomes, people typically exhibit negative time preference (i.e., they prefer an improving series of events). In the current study college students responded to hypothetical questions via an online survey. Part 1 was a replication and extension of procedures described by Loewenstein and Prelec (1991, section II). Response patterns like those of Loewenstein and Prelec were observed, in that the percentage of participants who saved the best for last decreased when the interval between activities in the sequence increased. In Part 2 participants were surveyed about their preference for the order in which they would experience hypothetical outcomes with sequences of different sizes (e.g., 3 activities to sequence or 8). These questions were answered: How does array size influence saving the best for last? How does array size influence preference for a perfectly improving sequence? As array size increased, the percentage of participants who saved the best for last, or generated a perfectly improving sequence, decreased. Regardless of array size, saving the best for last appears to be a more robust phenomenon than preference for perfectly improving sequences.

Key words: preference for sequences, time, array, delay, save the best for last, negative time preference

Save the Best for Last I: Young Adults Demonstrate Negative Time Preference- A Replication and Extension

Delay discounting research has shown that in choice paradigms where only one outcome is selected from the array, delayed outcomes are generally valued less than those without a delay (Ainslie, 1975). In keeping with this logic, if one were presented with an array of five items and given a chance to access each in a sequence of one's own choosing, one might conclude that the item selected first would be the most preferred. Some readers will recognize this arrangement as that characterized by the multiple-stimulus without replacement (MSWO) preference assessment described by DeLeon and Iwata (1996). It is possible, however, that the most preferred item may not be correctly identified in an MSWO if individuals choose to save the best for last. Not selecting the highest preferred item may be a concern because lesser preferred items may not engender the same level of performance as more preferred items (Glover, Roane, Kadey, & Grow, 2008; Koehler, Iwata, Roscoe, Rolider, & O'Steen, 2005). There is some scattered evidence to suggest that this phenomenon may occur, as it has been reported in the literature (Soldberg, Hanley, Layer, & Ingvarsson, 2007), in numerous posters presented at conferences (including Becerra & Fahmie, 2014; Litchmore, Ivy, & Weaver, 2014; Pendharkar, Bourret, Nuzzolilli, & Upshaw, 2017; Roath & Fritz, 2015), and in our own clinical practice. Though the authors recognize that posters and clinical anecdotes do not parallel the rigor of peer review, they do suggest that a phenomenon may exist, which should prompt further investigation. For example, in clinical medicine, accumulated reports (Baum, Holtz, Bookstein, & Klein, 1973; Schenken, 1976) inspired the development of a large case-control study which confirmed a strong association between long-term use of certain high-dose contraceptives and a rare and potentially deadly tumor (Rooks et al., 1979). To date, there are no published explicit studies of the “saving the best for last” phenomenon as it relates to the MSWO arrangement. In fact, this may represent a textbook exemplar of the “file drawer” problem (e.g., Shadish, Zelinsky, Vevea, & Kratochwill, 2016); thus, the prevalence of the phenomenon is unclear.

The present study is the first of a series of now completed studies conducted with the protracted goal of determining the extent to which saving the best for last occurs in the MSWO (DeLeon & Iwata, 1996). Before asking this question directly, a series of preliminary studies was conducted to determine the prevalence of this phenomenon, and to explore some factors that may contribute to saving the best for last. In the present submission, we investigated how the timing of choices and the number of outcomes in a sequence influenced participants’ choice to save the best for last. In a concurrent submission, we investigated how categorical features of sequences influence saving the best for last by college students and typically-developing preschoolers. Further, we investigated correspondence in college students’ preference for sequences involving hypothetical or real outcomes. Having laid forth the aims of this series of studies, we turn next to relevant research related to saving the best for last.

Traditionally, studies of intertemporal choice (i.e., “decisions in which the moment of choice and the associated consequences are separated in time;” Lucci, 2013, p. 1) have involved the delay-discounting model. This model is concerned with how people evaluate or select single outcomes from an array of multiple possible outcomes, at a point in time (see Frederick, Loewenstein, & O’Donoghue, 2002, for a history on the discounted utility model). Although the rate of discounting varies within and across individuals, studies on delay discounting using considerably diverse procedures have suggested that as delays increase, the value of an outcome diminishes according to a hyperbolic function (Madden & Johnson, 2010). In other words, for intertemporal choice involving single-outcome prospects (i.e., only one outcome is selected from the array), delayed outcomes are generally valued less than those that are not delayed (across species, and within humans of different characteristics; Ainslie, 1975). In economics, this is referred to as *positive time preference* (Olson & Bailey, 1981).

If positive time preference (i.e., the more “impulsive” choice) is evinced in all of one’s choices, then that would imply that when asked to schedule a set of outcomes, a person should typically prefer a worsening sequence (i.e., choose to start with the best outcome, followed by the second best, and ending with the worst outcome). Several studies have specifically focused on this problem and have shown that when a choice is viewed as part of a sequence of outcomes, people typically exhibit *negative time preference* (i.e., prefer an improving series of events; Loewenstein & Prelec, 1991; Magen, Dweck, & Gross, 2008). *Negative time preference* refers to a general preference for improvement, which can manifest in a perfectly improving sequence or in a sequence that simply ends with the most preferred option.

Taken together, the existing research shows that the framing of the choice affects how people behave (Tversky & Kahneman, 1981). More specifically, when the frame draws attention to individual components of the choice (e.g., would you like to read *or* watch a television show now?), people tend to exhibit positive time preference (Frederick et al., 2002), choosing their most preferred immediately; when the decision frame draws attention to the sequential aspect of the choice (e.g., which would you like to do *first*:

read or watch a television show?), people tend to exhibit negative time preference, saving their most preferred for last, in a sense "delaying gratification" (Loewenstein & Prelec, 1991; Magen et al., 2008).

With a brief three-question survey, Loewenstein and Prelec (1991, 1993) investigated how manipulating a specific parameter of a sequence, namely the timing of the outcomes, influenced people's choices in a predictable manner (i.e., showing positive or negative time preference). In the first question, the series of outcomes unfold over a fairly short period (2 weeks), with a brief (1 week) interval between the possible outcomes. In this case, 90% of subjects showed negative time preference, choosing to save the best for last. In the second question, there was a longer interval between both outcomes (i.e., the first outcome would be experienced immediately, whereas the second one would be experienced after a 6-month delay). Therefore, participants could potentially view this choice as two independent single-outcome choices instead of a sequence. Indeed, a significantly smaller percentage of participants (52%) demonstrated negative time preference given the long absolute delay. In the third question, both outcomes were delayed about 6 months, but the sequence interval (i.e., the timing between the outcomes) was once again reduced to 1 week. Like in the first question, most participants (83%) exhibited negative time preference. These results suggest that when outcomes are presented as a sequence, negative time preference is most likely, but the percentage of participants who prefer an improving sequence decreases when the interval between outcomes is increased.

Studies involving different methods (e.g., differences in the number of outcomes in the sequence, whether outcomes were hypothetical or real, or whether participants' ratings involve prospective evaluations before they experience the sequence or retrospective evaluations of sequences they have already experienced), have documented preference for improving sequences across a wide variety of outcomes including money (e.g., Duffy, Smith, & Woods, 2015; Magen et al., 2008), pain (e.g., Ariely & Carmon, 2000), annoying sounds (e.g., Ariely & Zauberman, 2000), restaurant meals (e.g., Loewenstein & Prelec, 1991, 1993), health outcomes (e.g., Chapman, 1996, 2000), affective stimuli (e.g., Drolet, Lau-Gesk, & Scott, 2011), and news (e.g., Legg & Sweeny, 2014; Marshall & Kidd, 1981).

In addition to showing that the timing or spacing of outcomes influences people's choices predictably, prior research has also suggested that if presented with more than one outcome of the same valence (i.e., perceived value), people prefer to spread such outcomes over time rather than concentrating them (Loewenstein & Prelec, 1993). For example, when deciding the order of four dinners, two at home and two at a fancy restaurant, people may prefer to intersperse dinners at home with dinners at a fancy restaurant, rather than accumulating the dinners at the fancy restaurant in one extreme of the sequence. Because Loewenstein and Prelec (1993) compared five-outcome sequences with only three distinct valences (i.e., two low preferred, two moderately preferred, and one highly preferred) whether sequences involving outcomes of distinct valence would result in similar preferences for overall improvement and spreading is a matter not yet resolved (Chapman, 1996; Guyse, Keller, & Eppel, 2002). Further, it is unclear whether distinct preference patterns would emerge with sequences of different array sizes (i.e., different total number of outcomes).

Evaluations of meaningful parameters, such as sequence size or category of outcomes, are needed to determine the prevalence of "saving the best for last" and preference for improving sequences. Although not explicitly compared yet, the number of outcomes in a sequence may influence people's choices for the temporal ordering of the sequence. Studies on preference for sequences of affective events have found preference for improving sequences when they involved only three outcomes (Drolet, Lau-Gesk, & Scott, 2011), but preference for mixed sequences when sequences involved 30 outcomes (Löckenhoff, Reed, & Maresca, 2012). Conceivably, preference for improvement is most probable given smaller sequences, and mixing preferred with less preferred outcomes is more probable for larger sequences. This conclusion must be stated speculatively, however, because the methods in the two cited studies differed in more ways than just the number of outcomes in the sequences.

Furthermore, some studies explicitly evaluating methodological variations have found that the preference for improvement may not be as robust as suggested by the extant literature on preferences for sequences (Frederick & Loewenstein, 2008). For example, Frederick and Loewenstein (2008) found that when the task involved creating sequences by allocating a fixed quantity of goods over a series of time periods (e.g., schedule 20 events over a 5-year period), respondents allocated outcomes evenly about half of the time. The rest of the time, improving or worsening sequences were equally likely.

Given the replication crisis in the social sciences, independent support for or against a set of findings, such as preference for improving sequences, should be of considerable value to the research community. An analysis of replications of 100 experiments reported in papers published in 2008 in three high-ranking psychology journals revealed that in nearly 62% of the replication attempts the results were not significant, when they had been in the original studies (Open Science Collaboration, 2015). Concerns with the replicability of findings is not limited to the social sciences. Behavior analysts have also started to express concern about publication bias (Tincani & Travers, 2019), and have called for more systematic replications and publications of failures to replicate (Hanley, 2017; Hantula, 2019; Perone, 2019). Doing so may be necessary if the goal is a thoroughgoing account of human experience.

Method

Participants and Setting

A total of 279 undergraduate college students participated in this study (see Table 1 for demographic information). Participants were recruited from psychology courses and invited to participate in the study via a psychology participant pool, for extra credit. The study was completed over a nine-month period, from March through December.

Measures and Procedures

Participants responded to an online survey in Qualtrics®XM (Qualtrics, Provo, UT).

Part 1. Time between events and sequence duration as predictors of positive time preference. The first three questions of Part 1 involved a *systematic*, rather than a *direct*, replication (Sidman, 1960) of procedures described by Loewenstein and Prelec (1991, 1993). Although the first three questions are identical to those of Loewenstein and Prelec, we used a different method (i.e., Qualtrics®XM) to survey participants. Moreover, participant recruitment differed in both studies. Specifically, we recruited undergraduate students from a university in the mid-Atlantic region and Loewenstein and Prelec recruited visitors to the Museum of Science and Industry in Chicago.

Table 1

Participant Characteristics

Demographic Characteristics	*f* (%)
Self-identified gender	
Man	55 (19.71)
Woman	223 (79.93)
Gender variant/non-conforming	1 (0.36)
Age	
18-24 years old	237 (84.95)
25-34 years old	27 (9.68)
35-44 years old	10 (3.58)
45-54 years old	5 (1.79)
Major	
Psychology	118 (42.29)
Biological sciences	60 (21.51)
Other	94 (33.69)
No response	7 (2.51)

Note. *f* = frequency of participants.

The first page of the survey consisted of the same three questions used by Loewenstein and Prelec. The scenario and questions appeared on the screen at the same time and are described next.

> Imagine you must schedule two weekend outings to a city where you once lived. You do not plan on visiting the city after these two outings.
>
> You must spend one of these weekends with an irritating, abrasive aunt who is a horrendous cook. The other weekend will be spent visiting former work associates whom you like a lot. From the following pairs, please indicate your preference by checking the appropriate line.
>
> (1) Suppose one outing will take place this coming weekend, the other the weekend after.
> - o This weekend **friends**; next weekend **abrasive aunt**
> - o This weekend **abrasive aunt**; next weekend **friends**
>
> (2) Suppose one outing will take place this coming weekend, the other in 6 months (26 weeks).
> - o This weekend **friends**; 26 weeks from now **abrasive aunt**
> - o This weekend **abrasive aunt**; 26 weeks from now **friends**
>
> (3) Suppose one outing will take place in 6 months (26 weeks from now), the other the weekend after (27 weeks from now).
> - o 26 weeks from now **friends**; 27 weeks from now **abrasive aunt**
> - o 26 weeks from now **abrasive aunt**; 27 weeks from now **friends**

The second page of the survey, described next, is an extension of Loewenstein and Prelec (1991, 1993) and included three similar questions modified slightly to reflect smaller time windows (i.e., within the same hour, same day, or a day apart).

> Imagine you are visiting a city where you once lived. You do not plan on visiting the city after this trip. You must schedule two outings during your trip. You must spend one of these outings with an irritating, abrasive aunt who is a horrendous cook. The other outing will be spent visiting former work

associates whom you like a lot. From the following pairs, please indicate your preference by checking the appropriate line.

(4) Suppose one outing will take place from 10:00 to 10:25, the other 10:30 to 10:55.

- 10:00 – 10:25 **friends**; 10:30 – 10:55 **abrasive aunt**
- 10:00 – 10:25 **abrasive aunt**; 10:30 – 10:55 **friends**

(5) Suppose one outing will take place at 9:00 am, the other at 1:00 pm on the same day.

- 9:00am **friends**; 1:00pm **abrasive aunt**
- 9:00am **abrasive aunt**; 1:00pm **friends**

(6) Suppose one outing will take place on the 21^{st} of the month, the other on the 22^{nd} of the same month.

- 21^{st} of the month **friends**; 22^{nd} of the month **abrasive aunt**
- 21^{st} of the month **abrasive aunt**; 22^{nd} of the month **friends**

Part 2. Number of items in the sequence as a predictor of positive time preference. The second part of the survey was a variation of the survey used in Study 1 of Loewenstein and Prelec (1993) and consisted of questions in which participants selected the order in which to experience the hypothetical outcomes in sequences of different sizes. Participants were first provided with this instruction:

> For the second part of the survey, you will need to think of brief activities that you like to different degrees and can be finished within a day.
>
> Think of something you really enjoy doing with people you really like, a favorite activity.
>
> Think of things you enjoy, but do not love, doing.
>
> Think of something you do not enjoy doing, a boring and not at all fun activity.

> In the space below, list 8 activities that you like to different degrees. Rank them from most preferred (1) to least preferred (8).

The activities listed by the participant were presented as a reminder throughout the rest of the survey. For the first 153 participants, the sequence questions were presented in an increasing order (starting with two outcomes and ending with eight). To minimize potential order effects (Chapman, 2000), the second 126 participants were randomly assigned to one of three conditions in which the array sizes were presented in three possible orders: an increasing order (defined above), a decreasing order (starting with eight and ending with two), or one randomly generated order (experiencing the following order of array sizes: 7, 4, 8, 3, 5, 6, 2). Each sequence consisted of outcomes of similar duration that could all be experienced within the same day without regard to scheduling issues (i.e., the participant could experience the outcome whenever chosen, and should therefore ignore scheduling considerations).

As in Study 1 of Loewenstein and Prelec (1993), the arrays consisted of boxes of different heights, which represented the participant’s preference for different leisure activities. A sample diagram with a randomly generated order was provided, and the participants selected the order in which they would like to experience each activity. An example of the instructions for a three-item array is presented below:

Imagine it is early morning and you are contemplating how to schedule your day. Your situation is depicted in the diagram below (see Figure 1).

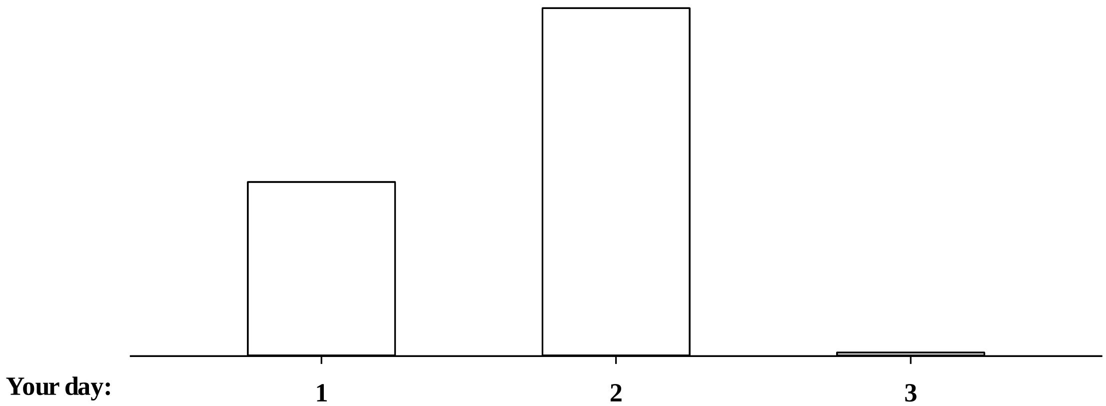

Figure 1. Sample diagram of a three-item array included in the instructions to participants.

The height of the box represents how much you enjoy the activity. The taller the box, the more you like that activity. In the diagram, you are doing the most pleasurable thing second.

The medium height box represents a moderately pleasurable activity. Think of something you enjoy, but do not love, doing. In the diagram, you are doing the moderately pleasurable activity first.

The flat line represents a boring and not at all fun activity. Think of something you do not enjoy doing. In the diagram, you are experiencing the most boring activity last.

As a reference, here is the list of activities you previously generated, in your preferred order: [the eight activities that had been provided at the beginning of Part 2, would appear hear].

Imagine it is early in the morning, how would you like to schedule your day? Select the order in which you would like to experience each activity (first, second, last).

The answer section consisted of a matrix, in which the different boxes were presented as column headers, and the order options (i.e., first, second, …last) as line headers. The participants selected the radio buttons within the matrix that corresponded to their choice.

Results and Discussion

Part 1

How does the timing of sequences influence saving the best for last?

Table 2 includes the data on the direct replication and systematic extension of Loewenstein and Prelec (1991, 1993), presented as the percentage of participants who selected to see their abrasive aunt first, and their friends last for all six questions in Part 1.

Table 2

Percentage of Participants who Selected to Save the Best for Last in Each Question of Part 1

Time frame in choice	Percentage who Saved the Best for Last
1. This coming weekend and the weekend after.	80
2. This coming weekend and in 6 months (26 weeks).	37*
3. In 6 months (26 weeks from now) and the weekend after (27 weeks from now).	68*
4. From 10:00 to 10:25 and 10:30 to 10:55.	82
5. At 9:00 am and at 1:00 pm on the same day.	86
6. On the 21^{st} of the month and on the 22^{nd} of the same month.	81

Note. The table displays the data on the direct replication (lines 1-3) and systematic extension (lines 4-5) of Loewenstein and Prelec (1991, 1993).

* Significantly different from that obtained by Loewenstein and Prelec (1991, 1993), $p < .05$.

For almost all the time windows, the majority (range, 68% to 86%) chose to schedule the outing with their friends last in the sequence. The only exception was the second question, in which the interval between both outings was much greater (6 months), and only 37% of participants exhibited negative time preference. It is possible that all other variables are relatively equal when there is a short interval within outcomes, but other variables (e.g., weather, time of year, professional responsibilities) could influence choice when there is a greater interval between outcomes. Nevertheless, Loewenstein and Prelec hypothesized that negative time preference is evinced in choices in which outcomes are *perceived* as part of a meaningful sequence. By increasing the interval between the outcomes to six months, the choice may be viewed as selecting the timing of two distinct outcomes rather than a whole sequence.

We first discuss the results for the first three questions, the replication of Loewenstein and Prelec (1991, 1993), and then the second three questions, the extension. The patterns observed in our results in the first three questions are like those reported by Loewenstein and Prelec, in that the highest percentage of participants who showed negative time preference was observed for question 1, when the outcomes occurred sooner and unfolded over a fairly brief period (1-week interval). The second highest percentage was observed for question 3, in which the interval outcome was also brief (1 week), but both were delayed by six months (26 weeks and 27 weeks). The smallest percentage was observed for question 2, in which the absolute interval between both outings was 6 months apart (this weekend and 26 weeks from now). The researchers conducted a chi-squared goodness of fit test to compare the percentage of participants who showed negative time preference in each of the first three scenarios to those same percentages reported by Loewenstein and Prelec. A similar percentage of participants in both samples displayed negative time preference in question 1, when the outing options were this weekend or next weekend (i.e., differences in the percentages were not statistically significant: χ^2 (1) = 2.704, p = .1). In the present sample, however, a significantly smaller percentage of participants chose to save the best for last in question 2, when the outings were spaced six months apart (χ^2 (1) = 3.854, p = .049), and in question 3, when the interval was again reduced to one week, but both outings were delayed by six months (χ^2 (1) = 4.391, p = .036).

Although the patterns were similar across the two studies, significantly fewer participants in our study chose to save the best for last when delays were involved (i.e., second and third questions), as compared to that of Loewenstein and Prelec (1991, 1993). For the last three questions in Part 1, which involved smaller time windows, a large percentage of participants, chose to save the best for last, just like they did for the first question.

Part 2

In Part 2, participants experienced the change in array size in one of three different orders: increasing order (starting with two outcomes and ending with eight), decreasing order (starting with eight and ending with two), and one randomly generated order (experiencing the following order of array sizes: 7, 4, 8, 3, 5, 6, 2). Results of Part 2 of the study are presented next. When the differences across groups were not statistically significant, the results were combined into a single group. When differences observed across groups, results were described per group.

How does array size influence saving the best for last? Figure 2 shows the percentage of participants who saved the best for last (i.e., who chose to put the tallest box last in the sequence) for each array size.

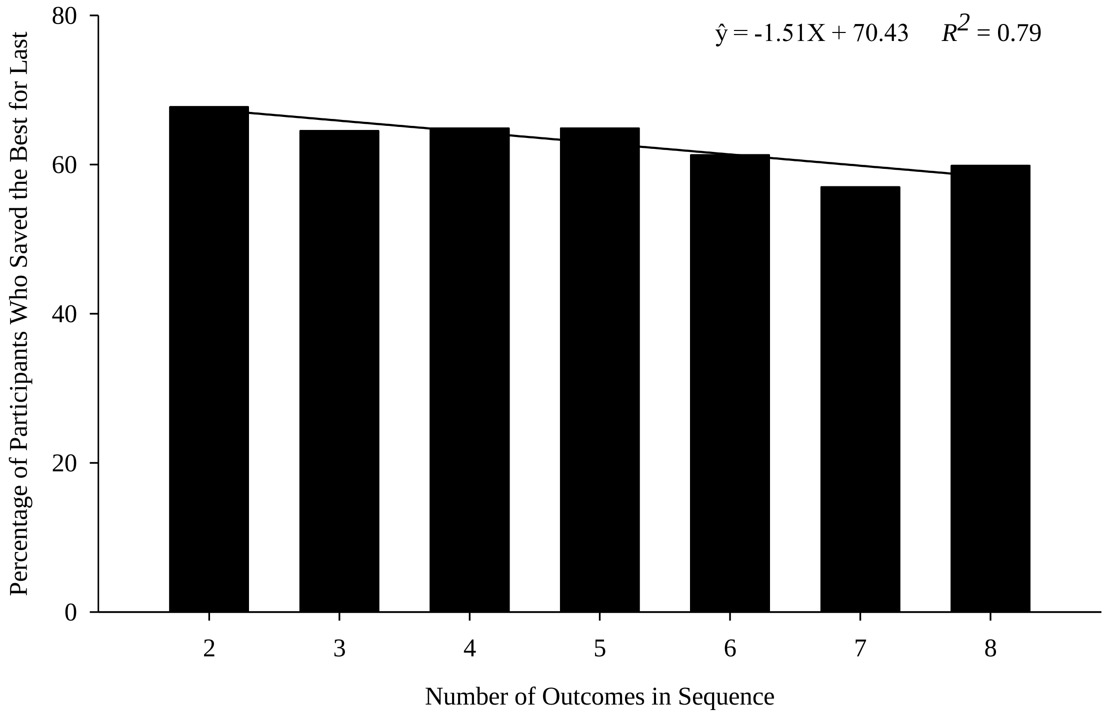

Figure 2. Percentage of participants who saved the best for last for each array size in Part 2. The percentage of participants who saved the best for last decreased as the array size increased, as indicated by the negative slope of the solid black regression line (B = -1.51, p = .007).

Combining all array sizes, more than half of participants M = 62.88%, range, 56.99% to 67.74%) of participants chose to save the best for last. The experimenters conducted a linear regression analysis to determine how well array size could predict the percentage of participants who saved the best for last. Results of the regression analysis indicated that the effect of array size on the percentage of participants who saved the best for last was significant (F (1, 5) = 19.08, p = .007). The regression equation for predicting the percentage of participants who saved the best for last using array size is $\hat{Y} = -1.51X.$ $+ 70.43$ and resulted in a R^2 of .79. That is, 79% of the variance in percentage of participants who saved the best for last was predicted from array size. The negative slope of the linear regression (B = -1.51, $p = .007$, 95% CI [-2.39, -.62]) indicates that as array size increased, the percentage of participants who saved the best for last decreased.

How does array size influence preference for a perfectly improving sequence? Does the order of exposure matter? To avoid redundant data, results for the array size of two are not included in these analyses, because saving the best for last is equivalent to generating a perfectly improving sequence when there are only two outcomes in the sequence. Table 3 shows the percentage of participants who generated a perfectly improving sequence in each group, for array sizes 3 through 8.

For each array size, we conducted chi-squared goodness of fit tests to compare the percentage of participants who generated perfectly improving sequences in each group, and results are depicted in the last columns of Table 3. The percentage of participants who generated perfectly improving sequences was not statistically different across order groups for array sizes 3 through 5. Chi-squared tests for arrays of six, seven, or eight outcomes suggested that the difference in percentages across groups was statistically significant ($p < .05$). Therefore, these data suggest that the effect of array sizes varied by the order of exposure. More specifically, the percentage of participants who generated a perfectly improving sequence for array sizes 6, 7, and 8, was greatest when those were presented toward the end of the survey. In other words, individuals were more likely to generate a perfectly improving sequence in the larger array sizes when those larger arrays were experienced at the end of the survey, as was the case for the increasing group.

Table 3

Percentage of Participants who Generated a Perfectly Improving Sequence for Array Sizes 3 through 8 Separated by Experience-Order Group

	Order				
Array Size	Increasing	Decreasing	Random	χ^2 (df=2)	*p*
3	51.0	42.9	39.5	2.393	.302
4	44.3	35.7	27.9	4.410	.110
5	40.2	23.8	37.2	3.970	.137
6*	40.7	21.4	27.9	6.981	.030
7*	39.7	19.0	14.0	14.750	.001
8*	37.6	14.3	20.9	11.363	.003
n	194	42	43		
$\hat{Y}$	-2.30X +54.90	-5.58X +56.87	-4.12X +50.57		
R^2	.797	.924	.639		

Note. χ^2 (df = 2) = Chi-squared tests comparing the percentage of participants who generated a perfectly improving sequence across groups. *p* = test of significance. Percentage of participants who generated a perfectly improving sequence did not differ significantly across groups for array sizes 3 through 5, but did differ significantly for array sizes 6 through 8, as indicated by the asterisk (*). *n* = sample size. $\hat{Y}$ = regression equation. R^2 = variance explained by the array size.

For each experience order group, the experimenters conducted linear regression analyses with percentage of participants who generated a perfectly improving sequence as the dependent variable and array size as the predictor. For this analysis, we also excluded data for the array of two outcomes. Results of the regression analysis indicated that the effect of array size on the percentage of participants who generated a perfectly improving sequence was significant for the groups that experienced the array sizes in increasing ($F(1, 4) = 15.74$, $p = .017$) and decreasing order ($F(1, 4) = 48.37$, $p = .002$), but was not significant for the group that experienced the random order ($F(1, 4) = 7.07$, $p = .056$). The regression equations for predicting the percentage of participants who generated a perfectly improving sequence using array size, for each group are depicted in Table 3. The R^2 for these equations for the increasing, decreasing, and random group indicate that 79.7% of the variance in percentage of participants who generated a perfectly improving sequence was predicted from array size for the increasing group, 92.4% for the decreasing group, and only 63.9% for the random group. The negative slope of the linear regressions for the increasing (B = -2.29, $p = .017$, 95% CI [-3.91, -.69]) and decreasing (B = -5.58, $p = .002$, 95% CI [-7.81, -3.35]) groups indicate that as array size increased, the percentage of participants who generated a perfectly improving sequence decreased.

Table 4 summarizes the percentage of all participants who saved the best for last, as well as the two most common sequence patterns generated for each array size, namely perfectly improving and perfectly worsening sequences. In all array sizes, the most commonly generated sequence was a perfectly improving sequence, indicative of negative time preference. The second most common sequence was a perfectly worsening sequence, indicative of positive time preference (i.e., delay discounting). For the array size of 3, the third most commonly generated sequence (16.49%) involved scheduling the moderately preferred outcome first, followed by the least preferred, and ending in the most preferred. All other sequences, across all array sizes, were generated by fewer than 9% of the sample, therefore, they will not be described specifically, but can be made available upon request.

Table 4

Frequency (and Percentage) of Participants Who Saved the Best for Last or Generated the Two Most Common Sequence Patterns Per Array Size in Part 2

Array Size	Saved the Best for Last	Perfectly Improving	Perfectly Worsening
2	189 (67.74)	189 (67.74)	90 (32.26)
3	180 (64.52)	134 (48.03)	55 (19.71)
4	181 (64.87)	113 (40.50)	50 (17.92)
5	181 (64.87)	104 (37.28)	48 (17.20)
6	171 (61.29)	100 (35.84)	51 (18.28)
7	159 (56.99)	91 (32.62)	49 (17.56)
8	167 (59.86)	88 (31.54)	48 (17.20)

Note. Participants who generated a perfectly improving sequence also saved the best for last, but some participants who saved the best for last did not generate perfectly improving sequences for array sizes 3 through 8.

Looking at array sizes 3 – 8, for the entire sample combined, a mean of 37.63% (range, 31.54% to 48.03%) of participants generated a perfectly improving sequence, demonstrating negative time preference. A smaller percentage of participants generated a perfectly improving sequence for all array sizes compared to the percentage of participants who saved the best for last, regardless of the order in which it was experienced. With the percentages ranging from 31.54% for the sequence of eight outcomes, to 48.03% for the sequence of three outcomes, both percentages were lower than the smallest percentage who saved the best for last (56.99% for the sequence of seven outcomes). Chi-squared tests comparing the percentages for each array size indicated that the differences were all statistically significant ($p < .001$). Taken together, approximately 62% of participants arranged selections to save the highest ranked item for last, whereas a smaller subset of those individuals also arranged all selections to progress from “worst to best” (37.63%).

General Discussion

The current set of studies aimed to investigate whether the timing of choices and the number of outcomes in a sequence influence participants’ choice to save the best for last. In Part 1, we presented college students with the same questions that Loewenstein and Prelec (1991, 1993) used and a variation of those questions. In the current study, observed response patterns were like those observed by Loewenstein and Prelec and by Andrade and Hackenberg (2012), when using questionnaires. Specifically, the percentage of participants who exhibited negative time preference was highest when both options occurred relatively soon and with a 1-week interval between options, the second highest percentage was observed when the interval remained the same but both options were delayed by six months, and the lowest percentage was observed when the interval between options was increased to six months. There was, however, a significant reduction in the percentage of participants in the current sample who exhibited negative time preference when there were delays involved, compared to those reported by the other two studies. One can only speculate as to why a smaller percentage of the current sample chose to save the best for last when there was a delay involved. Plausibly, differences in sample demographic characteristics account for the difference in responding. Perhaps there is something inherently different between undergraduate students in a university in the mid-Atlantic region and visitors to the Museum of Science and Industry in Chicago. It is possible that it is a cohort effect. Contemporary wisdom (Gleick, 2000) seems to suggest that, with technological advances, aspects of our society have, in a sense, sped up. Along with the internet, which became popularized in the early-to-mid 1990s with the invention of the world-wide web (Leiner et al., 2009), came substantive advances that have reduced the delay to many things like information, music, movies, mail, deliveries of online shopping, and communication. We have therefore become more sensitive to delays or more impatient (American Psychological Association, 2008). This sensitivity to delays in sequence outcomes is an area suitable for further direct and systematic replication.

One limitation of the current study is that the experimenters did not conduct a delay-sensitivity test, or a single-choice delay discounting question, such as “Would you rather spend time with your friends this weekend, or in two weekends?” This procedure was not included in the spirit of replicating the procedures described by Loewenstein and Prelec (1991, 1993) as closely as possible, however, this would be an ideal addition to a future replication.

Interestingly, the percentage of the current sample who saved the best for last in Part 2, when there were only two outcomes in the array, was smaller than it was in Part 1 (on all questions except number 2), which also involved sequences of only two outcomes. Methodological differences may account for these discrepancies. In Part 1, participants were provided a clear description of the different outings in the sequence. In Part 2, however, a list was not provided, and participants were prompted to think of brief activities that they like to varying degrees including favorite activities, things they enjoy but do not love doing, and boring activities they do not enjoy. Therefore, in Part 2, there was variability, within and across participants, in the type of outcomes that they ranked during the survey. For example, some participants included only items that could be considered leisure activities such as listening to music, spending time with friends, going on nature walks, drawing, and bowling. Whereas other participants included activities such as taking out the trash, doing chores, watching a bad movie, doing homework, going to the dentist, or going to a funeral. Perhaps there are qualitative aspects of the outcomes imagined that exerted control over their response. Furthermore, although there is precedent in the literature for using bars of different heights to represent relative quantitative or qualitative differences in options from which participants choose (Chapman, 2000, Löckenhoff, Rutt, Samanez-Larkin, O’Donoghue, & Reyna, 2019; Loewenstein & Prelec, 1993), it is possible that the boxes in the diagram may have been too abstract for some participants to visualize the hypothetical activities they were to schedule. A more circumscribed list of outcomes provided to participants to rank in Part 2 may have produced very different results.

Finally, the aforementioned discrepancy in results supports Andrade and Hackenberg’s (2012) conclusion that “different methods may occasion different decision-making tendencies” (p. 61). Although Andrade and Hackenberg replicated the previous findings of Loewenstein and Prelec (1991, 1993) when using questionnaires, they did not when participants were making repeated choices of sequences of (actual) video clips.

Thus, when participants were choosing among consumable reinforcers, the results were not consistent with negative time preference. As such, comparisons of participant responding when the sequence outcomes are hypothetical or real is an important area for future research.

Finally, the results of Part 2 suggest that "saving the best for last" may be a more robust phenomenon than a preference for perfect improvement. Although the percentage of participants who elected to save the best for last was high for all array sizes, that percentage of participants who generated a perfectly improving sequence was significantly lower. Furthermore, there was an interaction effect between the size of the array and the order of exposure, in that generating perfectly improving sequences for arrays of six, seven, or eight outcomes was more likely when those where experienced toward the end of the survey. The authors believe this is an area that is worth exploring further, as it may shed light on potential interventions for impulsive behavior. Perhaps, exposing people to sequences of increasing array sizes may promote negative time preference, self-control, or delay of gratification.

From a behavior analytic perspective, savoring the moment and anticipating the future may involve intrinsically reinforcing self-descriptive verbal behavior. Life is generally more pleasant when we have things to look forward to than when we do not. Booking a vacation is precurrent behavior necessary for anticipating and daydreaming about the forthcoming vacation. Engaging in anticipating and daydreaming are likely intrinsically reinforced activities. Likewise, to anticipate and savor the last bite of the most palatable dish on our plate, requires that it remains on the plate until the end. Engaging in this anticipatory behavior may have immediate value (i.e., reinforcing efficacy) that depends on the scheduling of the terminal reinforcer. Empirically backing up this interpretation of negative time preference would be difficult using methods that behavior analysts are traditionally comfortable using. However, if we accept self-report choice data, perhaps we would be open to accepting, albeit with some salt of skepticism, reports about why people make those choices.

Although the present studies involved a non-clinical population and hypothetical outcomes, determining the extent of the phenomenon (negative time preference or saving the best for last) was thought to be a foundational precursor to applied explorations of behavior change. The authors view self-reports as *one method* for studying human behavior. Recognizing the value of self-reports and expanding our methods in general may give our work as behavior analysts greater relevance beyond our field (Vyse, 2013). Furthermore, the authors believe these initial studies were necessary to determine whether the "saving the best for last" phenomenon is real and whether future research in this area should be pursued. The results of the current studies suggest that more than half of all participants (M = 62.88%; range, 56.99% to 67.74%) chose to save the best for last. Therefore, this phenomenon is, in our view, worthy of further study. In Part 2 of the current study, participants generated very diverse lists of outcomes, which could explain the reduced percentage of participants who saved the best for last, compared to Part 1. Consequently, we decided to study how categorically-different sequences (e.g., those involving school work, food, exercise, leisure, and noxious experiences) influence saving the best for last, as well as to compare correspondence of responses to hypothetical or real outcomes. These comparisons across outcome categories were conducted in a series of studies described in a concurrent submission.

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82*, 463-496. doi:10.1037/h0076860

American Psychological Association, Policy and Planning Board. (2008). *How technology changes everything (and nothing) in psychology: 2008 annual report of the APA Policy and Planning Board.* doi: 10.1037/a0015888

Andrade, L. F., & Hackenberg, T. D. (2012). Saving the best for last? A cross-species analysis of choices between reinforcer sequences. *Journal of the Experimental Analysis of Behavior, 98*, 45-64. doi: 10.1901/jeab.2012.98-45

Ariely, D., & Carmon, Z. (2000). Gestalt characteristics of experiences: The defining features of summarized events. *Journal of Behavioral Decision Making, 13*, 191-201. doi: 10.1002/(SICI)1099-0771(200004/06)13:2<191::AID-BDM330>3.0.CO;2-A

Ariely, D., & Zauberman, G. (2000). On the making of an experience: The effects of breaking and combining experiences on their overall evaluation. *Journal of Behavioral Decision Making, 13*, 219-232. doi: 10.1002/(SICI)1099-0771(200004/06)13:2<219::AID-BDM331>3.0.CO;2-P

Baum, J. K., Holtz, F., Bookstein, J. J., & Klein, E. W. (1973). Possible association between benign hepatomas and oral contraceptives. *The Lancet, 302*, 926-929. doi: 10.1016/s0140-6736(73)92594-4

Becerra, L., & Fahmie, T. A. (2014, May). *A review and analysis of the consistency of MSWO assessments.* Poster presented at the 40th Annual Convention of the Association for Behavior Analysis International, Chicago, IL.

Chapman, G. B. (1996). Expectations and preferences for sequences of health and money. *Organizational Behavior and Human Decision Processes*, *67*, 59-75. doi: 10.1006/obhd.1996.0065

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

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29*, 519-533. doi: 10.1901/jaba.1996.29-519

Drolet, A., Lau-Gesk, L., & Scott, C. (2011). The influence of aging on preferences for sequences of mixed affective events. *Journal of Behavioral Decision Making, 24*, 293-314. doi:10.1002/bdm.695

Duffy, S., Smith, J., & Woods, K. (2015). How does the preference for increasing payments depend on the size and source of the payments? *Management Science Letters, 5*, 1071-1080. doi: 10.5267/j.msl.2015.10.007

Frederick, S., & Loewenstein, G. (2008). Conflicting motives in evaluations of sequences. *Journal of Risk and Uncertainty, 37*, 221-235. doi:10.1007/s11166-008-9051-z

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, *40*, 351-401. doi:10.1257/002205102320161311

Gleick, J. (2000). *Faster: The acceleration of just about everything.* New York, NY: Vintage Books.

Glover, A. C., Roane, H. S., Kadey, H. J., & Grow, L. L. (2008). Preference for reinforcers under progressive- and fixed-ratio schedules: A comparison of single and concurrent arrangements. *Journal of Applied Behavior Analysis, 41*, 163–176. doi: 10.1901/jaba.2008.41-163

Guyse, J. L., Keller, L. R., & Eppel, T. (2002). Valuing environmental outcomes: Preferences for constant or improving sequences. *Organizational Behavior and Human Decision Processes, 87,* 253–277. doi: 10.1006/obhd.2001.2965

Hanley, G. P. (2017). Editor's note. *Journal of Applied Behavior Analysis, 50,* 3–7. doi: 10.1002/jaba.366

Hantula, D. A., (2019). Editorial: Replication and reliability in behavior science and behavior analysis: A call for a conversation. *Perspectives on Behavior Science, 42,* 1-11. doi: 10.1007/s40614-019-00194-2

Koehler, L. J., Iwata, B. A., Roscoe, E. M., Rolider, N. U., & O'Steen, L. E. (2005). Effects of stimulus variation on the reinforcing capability of nonpreferred stimuli. *Journal of Applied Behavior Analysis, 38*, 469–484. doi: 10.1901/jaba.2005.102-04

Legg, A. M., & Sweeny, K. (2014). Do you want the good news or the bad news first? The nature and consequences of news order preferences. *Personality and Social Psychology Bulletin, 40*, 279-288. doi:10.1177/0146167213509113

Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., … Wolff, S. (2009). A brief history of the internet. *Computer Communication Review, 39*, 22-31. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid&db=edsbl&AN=RN259600903&site=eds-live&scope=site

Litchmore, M., Ivy, J. W., & Weaver, M. T. (2014, May). *Reinforcer strength: An examination of assumptions of preference assessment results.* Poster presented at the 40th Annual Convention of the Association for Behavior Analysis International, Chicago, IL.

Löckenhoff, C. E., Reed, A. E., & Maresca, S. N. (2012). Who saves the best for last? Age differences in preferences for affective sequences. *Psychology and Aging, 27*, 840-848. doi:10.1037/a0028747

Löckenhoff, C. E., Rutt, J. L., Samanez-Larkin, G. R., O'Donoghue, T., & Reyna, V. F. (2019). Preferences for temporal sequences of real outcomes differ across domains but do not vary by age. *Journals of Gerontology: Psychological Sciences, 74,* 430-439. doi:10.1093/geronb/gbx094

Loewenstein, G., & Prelec, D. (1991). Negative time preference. *The American Economic Review, 81*, 347-352.

Loewenstein, G. F., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review, 100*, 91-108.

Lucci, C. R. (2013). Time, self, and intertemporal choice. *Frontiers in Neuroscience, 7,* 1-5. doi: 10.3389/fnins.2013.00040

Madden, G. J., & Johnson, P. S. (2010). A delay-discounting primer. In G. J. Madden & W. K. Bickel (Eds.), *Impulsivity: The behavioral and neurological science of discounting* (pp. 11-37). Washington, DC: American Psychological Association.

Magen, E., Dweck, C. S., & Gross, J. J. (2008). The hidden-zero effect: Representing a single choice as an extended sequence reduces impulsive choice. *Psychological Science, 19*, 648-649. doi:10.1111/j.1467-9280.2008.02137.x

Marshall, L. L., & Kidd, R. F. (1981). Good news or bad news first? *Social Behavior and Personality*, *9*, 223-226. doi:10.2224/sbp.1981.9.2.223

Olson, M., & Bailey, M. (1981). Positive time preference. *Journal of Political Economy, 89*, 1-25. Retrieved from http://www.jstor.org.proxy-bc.researchport.umd.edu/stable/1837349

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349, aac4716. doi:10.1126/science.aac4716

Pendharkar, T., Bourret, J. C., Nuzzolilli, A., & Upshaw, S. (2017, May). *Saving the best for last in MSWO preference assessments.* Poster presented at the 43rd Annual Convention of the Association for Behavior Analysis International, Denver, CO.

Perone, M. (2019). How I learned to stop worrying and love replication failures. *Perspectives on Behavior Science, 42,* 91-108. doi: 10.1007/s40614-018-0153-x

Qualtrics® XM (2018) [Online survey platform]. Provo, UT: Qualtrics.

Roath, C., & Fritz, J. N. (2015, May). *Order of selections in the Multiple-Stimulus Without Replacement preference assessment.* Symposium presented at the 41st Annual Convention of the Association for Behavior Analysis International, San Antonio, TX.

Rooks, J. B., Ory, H. W., Ishak, K. G., Strauss, L. T., Greenspan, J. R., Hill, A. P., & Tyler, C. W., Jr. (1979). Epidemiology of hepatocellular adenoma: The role of oral contraceptive use. *Journal of the American Medical Association, 242*, 644–648. doi: 10.1001/jama.1979.03300070040020

Schenken, J. R. (1976). Hepatocellular adenoma: Relationship to oral contraceptives? [Letter to the editor]. *Journal of the American Medical Association, 236*, 559. doi: 10.1001/jama.1976.03270060013003

Shadish, W. R., Zelinsky, N. A. M., Vevea, J. L., & Kratochwill, T. R. (2016). A survey of publication practices of single-case design researchers when treatments have small or large effects. *Journal of Applied Behavior Analysis, 49,* 656–673. doi: 10.1002/jaba.308

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology.* New York: Basic Books.

Soldberg, K. M., Hanley, G. P., Layer, S. A., & Ingvarsson, E. Y. (2007). The effects of reinforcer pairing and fading on preschoolers' snack selections. *Journal of Applied Behavior Analysis, 40*, 633-644. doi: 10.1901/jaba.2007.633–644

Tincani, M., & Travers, J. (2019). Replication research, publication bias, and applied behavior analysis. *Perspectives on Behavior Science, 42*, 59-75. doi: 10.1007/s40614-019-00191-5

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211,* 453-458. doi:10.1126/science.7455683

Vyse, S. (2013). Changing course. *The Behavior Analyst*, *36*, 123-135. doi: 10.1007/BF03392295

Save the Best for Last II: Whether One Saves the Best for Last Depends on Outcome Category

Mariana I. Castillo[1], Shuyan Sun[1], Michelle A. Frank-Crawford[1&3],

Griffin W. Rooker[2&3], and John C. Borrero[1]

UMBC[1]

Johns Hopkins University School of Medicine[2], and Kennedy Krieger Institute[3]

Completion of these studies and preparation of this manuscript were supported in part by grants from the UMBC Graduate Student Association.

Direct correspondence to J. C. Borrero, Department of Psychology, UMBC, 1000 Hilltop Circle, Baltimore, Maryland 21250 (e-mail: jborrero@umbc.edu).

Mariana Castillo ORCID 0000-0002-7433-4324

Shuyan Sun ORCID 0000-0002-7512-2753

Michelle Frank-Crawford ORCID 0000-0003-4319-6317

Griffin W. Rooker ORCID 0003-0246-0775

John Borrero ORCID 0000-0003-0413-0892

Abstract

Generally, people prefer immediate over delayed outcomes. As such, when arranging outcomes, one could assume a person would prefer to start with the best outcome and end with the worst outcome. Nevertheless, people typically exhibit negative time preference (i.e., they prefer an improving series of outcomes) when the choice involves a sequence of outcomes. The generality of this finding was assessed across types of stimuli and populations. In Study 1, 192 college students responded to hypothetical questions involving categorically-different outcomes (e.g., noxious stimuli, food) via an online survey. A significantly smaller percentage of participants saved the best for last relative to prior studies, but the percentage was highest when sequences involved noxious stimuli or food. In Study 2 we examined the correspondence between college students' preference for the order in which they would experience sequences of categorically-different outcomes when those were hypothetical versus real. There was very strong correspondence in the ranks assigned to the hypothetical and real outcomes, but more variability in the sequences generated. In Study 3 we aimed to determine preschoolers' preference for sequences. With academic items, 2 of the 4 participants chose to save the best for last. With leisure items, none of the participants saved the best for last. Preschoolers generally interspersed more- and less- preferred activities.

Keywords: negative time preference, positive time preference, preference for sequences, categories, save the best for last.

Save the Best for Last II: Whether One Saves the Best for Last

Depends on Outcome Category

Delay discounting studies have found that in choice paradigms where only one outcome is selected from an array, delayed outcomes are generally valued less than immediate ones (Ainslie, 1975). Specifically, *delay discounting* refers to a decrease in the value of a reward when it is delayed compared to when it is immediately available (Odum, 2011). Delay discounting studies often involve hypothetical choices of consumable or material commodities, such as money (Bialaszek, Ostaszewski, Green, & Myerson, 2019), food (Friedel, DeHart, Madden, & Odum, 2014), and legal (Bickel, Odum, & Madden, 1999) and illegal (Madden, Bickel, & Jacobs, 1999) substances. In economics, this diminished value of delayed outcomes relative to those that are not delayed, or alternatively, preference for present versus future consumption is referred to as *positive time preference* (Olson & Bailey, 1981).

There is evidence to suggest, however, that time preference may depend on the choice paradigm. Results of several studies have suggested that when a choice is viewed as being part of a sequence of outcomes, people typically exhibit *negative time preference* (i.e., they prefer an improving series of outcomes; Loewenstein & Prelec, 1991). Notably, *negative time preference* refers to a preference for future over immediate consumption of the more preferred option, or a general preference for improvement, which can manifest in a perfectly improving sequence or in a sequence that simply ends with the most preferred option. In the Loewenstein and Prelec (1991) study, for example, when the choice was between eating at a more preferred restaurant in one month or in two months, 80% of participants preferred the sooner option (i.e., they exhibited positive time preference). When the choice involved scheduling two dinners: one at a moderately preferred restaurant and one at a more preferred restaurant, 57% of participants elected to experience the moderately preferred meal in one month and the more preferred meal in two months (i.e., they exhibited negative time preference; Loewenstein & Prelec, 1991). The existing research shows that the framing of the choice affects how people behave (Tversky & Kahneman, 1981). More specifically, when the question frame draws attention to individual components of the choice (e.g., would you like to spend time with your friends now or later?), people tend to exhibit positive time preference (Frederick, Loewenstein, & O'Donoghue, 2002); when the decision frame draws attention to the sequential aspect of the choice, or a series of unavoidable outcomes (e.g., you are able to choose the order in which you complete your work assignment and spend time with friends; which would you like to do first?), people tend to exhibit negative time preference (Loewenstein & Prelec, 1991).

In addition to the framing of the choice, other factors are relevant to time preference. Two factors that have been demonstrated to have an impact on preference for sequencing are outcome category and real versus hypothetical outcomes. Outcome category refers to specific characteristics of the outcome (e.g., pleasurable or aversive outcomes, an experience or a material purchase); whereas real versus hypothetical refers to whether the outcomes are just imagined or truly experienced.

Studies have found that the effect of delays on time preference may depend on the category of the outcome—the type of outcome. In a set of experiments, Kumar, Killingsworth, and Gilovich (2014) found that participants derived pleasure from anticipating experiences (e.g., concerts, meals out, vacations), and waiting for experiences tended to be more enjoyable than waiting for material goods (e.g., clothing, gadgets). Although there is evidence that individuals sometimes prefer to delay a desirable outcome (Loewenstein, 1987, Kumar & Gilovich, 2016), preference for reducing the delay to aversive outcomes seems more widespread (Cook & Barnes, 1964; Story et al., 2013). When given the option to choose between an immediate or delayed inevitable shock, the majority (approximately 78%) of participants opted to receive more-painful shocks right away rather than wait for less painful shocks following a brief delay (Story et al., 2013). Story et al. posited that the anticipation of forthcoming and inevitable pain is so aversive in and of itself that people will pay a significant price, in the form of more physical pain, to avoid it.

Moreover, a study evaluating preferences for temporal sequences of categorically-different real experiences, found that participants seemed to prefer improving sequences for aversive physical experiences (i.e., decreasing intensities of shock) but mixed sequences for physical effort and monetary payouts (Löckenhoff, Rutt, Samanez-Larkin, O'Donoghue, & Reyna, 2019). Taken together, these results suggest that negative time preference in sequences is not universal, and that time preference may be influenced by the framing of the choice and by the category of outcomes involved.[i]

Though there is considerable evidence to support preference for categorically different outcomes, little is known about the correspondence between people's choices of real versus hypothetical *experiential* outcomes. Studies on delay discounting using real and hypothetical *monetary* rewards have found no significant differences in responding as a function of reward type (Lagorio & Madden, 2005; Locey, Jones, & Rachlin, 2011; Madden, Begotka, Raiff, & Kastern, 2003). Nevertheless, studies on correspondence between real and hypothetical choices using methods other than delay discounting tasks have found discrepancies when choices involve moral decisions (FeldmanHall et al., 2012), consumer goods purchases (Kang, Rangel, Camus, & Camerer, 2011), or strategic social interactions (Vlaev, 2012). Moreover, recent studies have found differences in risk-taking behavior and neurological activity when the rewards involved large magnitudes of hypothetical or real money (Xu et al., 2018), with participants taking less risks with increased magnitudes of real money, but not changing their risk-taking behavior with increased magnitudes of hypothetical money. Therefore, there are contradictory findings regarding correspondence between real and hypothetical choices, and this difference may be more pronounced when the outcomes involve things other than small amounts of money. We do not know if preference for sequences of categorically-different outcomes will remain constant when the outcomes are real versus hypothetical.

Thus far, the discussion has revolved around contextual factors that have an impact on choice. Nevertheless, individual characteristics may also influence choice and time preference as it relates to sequences[ii]. Among such characteristics, two are relevant to the current investigation, and have received limited attention as they relate to preference for sequences: impulsiveness and age.

Although time preference and impulsiveness are different constructs, the authors hypothesize that they may be related. Delay discounting, an example of positive time preference, is considered a measure of impulsivity (Odum, 2011), and discounting has been positively associated with self-report measures of impulsiveness (de Wit, Flory, Acheson, McCloskey, & Manuck, 2007). Therefore, it is possible that negative time preference related to the sequence frame is also associated with impulsiveness and self-control. Presumably, highly impulsive individuals may be less likely to exhibit negative time preference, or less likely to save the best for last, relative to less impulsive individuals.

Finally, there are no known published studies of preference for sequences with children. To date, most published studies have evaluated preference for sequences with young adults, with only a few looking at this phenomenon with older adults (e.g., Löckenhoff et al., 2019). Age differences have been observed in delay discounting, with children discounting future rewards at a greater rate than adults (Green, Fry, & Myerson, 1994). If sensitivity to delay changes throughout the lifespan, it is possible that preference for sequences, specifically for delaying gratification by generating improving sequences, or saving the best for last, may also strengthen with development. Given the lack of research with children in this area, we do not know whether preschool children will also exhibit negative time preference when asked to schedule sequences of outcomes.

Three studies were designed to answer the following questions: (a) does the likelihood of saving the best for last differ when the sequences involve noxious experiences, food, leisure items, schoolwork, exercise, or a mix of all categories? (b) is impulsiveness as measured by the BIS-11, and specifically, non-planning impulsiveness, related to time preference in sequences of hypothetical outcomes? (c) is there correspondence between college students’ preference for sequence order when outcomes are hypothetical versus real? and (d) will typically-developing preschool children also choose to save the best for last when scheduling sequences of school tasks, food, exercise, stories, leisure activities, or a mix of all categories?

Study 1: Hypothetical Outcomes

The purpose of Study 1 was to investigate the likelihood of finding negative time preference, and saving the best for last, in sequences of categorically-different hypothetical outcomes and their relation to self-report measures of impulsiveness. Specifically, we aimed to determine whether college students' likelihood of saving the best for last differed based on the category of the outcomes in the sequence (categories are explicitly described below). We also aimed to determine whether impulsiveness, and specifically non-planning impulsiveness (i.e., a lack of "future thinking"), is related to time preference in sequences. A thoroughgoing understanding of the roles outcome categories have on time preference could have implications for a wide range of potential applications: (a) promoting pro-health behavior (e.g., exercise), (b) facilitating tolerance or compliance with painful medical procedures, (c) enhancing employee performance in organizational settings, (d) improving educational outcomes, and (e) identifying preferred items (and potential reinforcers) for behavioral interventions for persons with intellectual and developmental disabilities.

Method

Participants and Setting

A total of 192 undergraduates participated in Study 1 (see Table 1 for demographic information). Participants were recruited from psychology courses and invited to complete an online survey via a psychology participant pool, for extra credit. Recruitment and data collection were completed over a two-month period, from mid-March through mid-May.

Table 1

Study 1 Participant Characteristics

Demographic Characteristics	*f* (%)
Self-identified gender	
Man	34 (17.71)
Woman	155 (80.73)
Gender variant/non-conforming	1 (0.52)
Not specified	2 (1.04)
Age	
18-24 years old	167 (86.98)
25-34 years old	19 (9.90)
35-44 years old	4 (2.08)
45-54 years old	2 (1.04)
Major	
Psychology	121 (63.02)
Biological sciences	32 (16.67)
Other	39 (20.31)

Note. N = 192. f = frequency of participants.

Measures and Procedures

Overview. Participants who contacted the research team were asked to respond to an online survey in Qualtrics®XM (Qualtrics, Provo, UT). The survey consisted of two parts completed consecutively in a single sitting: the Sequence Survey and the Barratt Impulsiveness Scale (BIS; Patton, Stanford, & Barratt, 1995). The impulsiveness scale was included to correlate with the results of the Sequence Survey and better understand the relation between time preference and impulsivity, as highly impulsive individuals may be less likely to exhibit negative time preference, or less likely to save the best for last, relative to less impulsive individuals.

Part 1: Sequence Survey. The survey was similar to that used in Study 2 of Castillo, Sun, Frank-Crawford, and Borrero (2020), except that in the current study verbal descriptions were used to represent the stimuli, rather than abstract graphical representations. The survey for the current study consisted of six sequences, each involving categorically-different outcomes (see Appendix for a text copy of the survey questions). Categories included: (a) noxious outcomes, (b) food, (c) leisure, (d) schoolwork, (e) exercise, and (f) a mixed category. Participants were first provided with a list of items in each category and asked to rank them in order of preference. Following informal interviews with undergraduate students, the study team members collectively generated a list for each category that included items with which most undergraduate students have presumably had experience. The food list included strawberries, broccoli, tomatoes, chips, and crackers. The leisure list included reading a magazine, watching YouTube®, spending time with their phones, playing a paper game (e.g., Sudoku or crossword puzzles), and playing a game on a computer or tablet. The school work list included three- and four-digit multiplication problems, long division, transcribing hand-written notes with a computer, reading comprehension, and building a model with plastic rods and connectors. The exercise list included jumping jacks, sit-ups, push-ups, running in place, and squats. Unlike the other categories, which included items with which most students have likely had experience, the list of noxious experiences was based on stimuli that are often used to study pain and sensory perception in clinical and laboratory settings (Yarnitsky & Pud, 2004). The items selected for the noxious category have been used extensively for neurological sensory testing with humans, including children with intellectual and developmental disabilities (Barney, Tervo, Wilcox, & Symons, 2017; Mitchell, MacDonald, & Brodie, 2004; von Baeyer, Piira, Chambers, Trapanotto, & Zeltzer, 2005). Specifically, noxious stimuli included submerging the hand in very cold water (approximately 1º C, or ~ 33º F), having the skin on the inner wrist pricked with a small plastic pin, experiencing the equivalent of approximately 1.8 kg of pressure on their inner wrist, having the end of a thick and stiff nylon thread pressed repeatedly against the skin on the inner wrist (30 times in 30 s), and having a warm probe (50º C, or 122º F) touch the skin of the inner wrist. The mixed category included the top-ranked item from each of the other categories.

After ranking the items in each categorical list (e.g., noxious outcomes 1-5), participants were asked to select the order in which they would like to experience the five different hypothetical outcomes in a sequential order. Category presentation was randomized across participants (some participants experienced the noxious category first, others the food category first, and so on), but the mixed category was always presented last.

An array of five outcomes was selected for individuals to sequence in a preferred order based on the results of Castillo et al. (2020). In that study, the sequence of five items was the largest array size in which the percentage of participants who generated a perfectly improving sequence did not differ based on the order in which it was presented in the survey (χ^2 (2) = 3.970, p = .137). Furthermore, when combining across experience order groups, the array size of five was significantly different from that of the sequence of three items (χ^2 (1) = 6.5779, p = .01).

Part 2: Barratt Impulsiveness Scale. Immediately following completion of the Sequence Survey, participants were asked to complete the BIS-11 (Patton et al., 1995) via Qualtrics. The BIS-11 is a 30 item self-report measure designed to assess impulsiveness (Patton et al., 1995). The BIS was originally published 60 years ago (Barratt, 1959), and at the time of its 50th anniversary, it was reportedly the most commonly administered self-report instrument for the assessment of impulsiveness in research and clinical settings (Stanford et al., 2009). Principal component analysis of the BIS-11 produced six first-order factors, and three second-order factors. These second-order subscales include Motor Impulsiveness (first-order factors motor and perseverance), Non-Planning Impulsiveness (first-order factors self-control and cognitive complexity), and Attentional Impulsiveness (first-order factors attention and cognitive instability). Self-report measures of self-control have been found to have high convergent validity (Duckworth & Kern, 2011), and scores on the non-planning subscale of the BIS-11 have been associated with delay discounting in previous studies (de Wit et al., 2007). In the current study, we conducted analyses using the overall score of the BIS-11, as well as the score in the Non-Planning Impulsiveness scale.

Results and Discussion

How do the categorical features of items in the array influence saving the best for last? To answer this question, we determined whether or not a participant selected the option they previously ranked #1 to be last in the sequence. Sequences consisted of five items. The probability of choosing the item ranked #1 in the last position, if participants were randomly clicking on the response options, was 0.2. Conversely, the probability of selecting one of the other four options last in the sequence was 0.8. We conducted a chi-squared test to determine whether the percentage of participants who saved the best for last in any given category was different from what would be expected had participants answered randomly. The results of this analysis are depicted in Table 2. In all categories, less than a third of the participants (range, 16.67% to 30.73%) chose to save the best for last in the sequence. The chi-squared test suggests that in only two categories, namely food and noxious stimuli, did the percentage of participants differ from what would be expected if participants had responded randomly. Specifically, when the sequences involved food or noxious stimuli, participants were more likely to save the best for last.

Table 2

Percentage of Participants who Selected to Save the Best for Last in Each Category of Study 1

Category	Percentage who Saved the Best for Last	χ^2 (1)	*p*
Food	30.73*	13.81	<.001
Noxious stimuli	28.65*	8.97	.003
Exercise	25.52	3.66	.060
School work	23.44	1.42	.230
Leisure	20.31	0.01	.910
Mixed	16.67	1.33	.250

Note. In only two categories (food and noxious stimuli), * the percentage of participants who saved the best for last was significantly greater than the expected percentage, had participants responded randomly ($p < .05$; exact *p*-values are reported in the *p* column).

How does impulsiveness correlate with the preference for sequences? For each sequence the researchers calculated a Spearman rank order correlation between the relative value of the outcome, determined by the rank assigned by the participant and the order in which it was selected. This correlation captures individual variations in sequence preferences. The resulting scores ranged from −1 (perfectly improving: rank is inversely related to position in sequence) to 1 (perfectly worsening: rank is directly related to position in sequence) with scores closer to zero indicating a preference for interspersing higher and lower preferred outcomes.[iii] A Spearman correlation coefficient was obtained for each category. Two Pearson correlations were calculated. The first correlation was between each participant's overall score in the BIS-11 and the coefficient of the Spearman rank-order correlation between preference rank and position order in the sequence. The second correlation was between each participant's score in the non-planning subscale of the BIS-11 and the Spearman rank-order correlation between preference rank and position order in the sequence. Table 3 depicts Cronbach's alpha reliability of the BIS-11 and the subscales, as well as the mean score and standard deviation for each.[iv] Table 4 depicts the Pearson's correlation coefficients between the Spearman rank-order correlation and the score in the entire BIS-11 or just the non-planning subscale. No significant associations were observed.

We conducted further exploratory analysis by separating the results into four different groups: (a) participants who generated diverse sequences but saved the best option to the end, (b) participants who generated perfectly improving sequences (i.e., worst-to-best), (c) participants who generated perfectly worsening sequences (i.e., best to worst), and (d) all other participants. Impulsiveness scores did not differ significantly across groups. These data are available upon request.

Table 3

BIS-11 and Subscale Reliability and Mean Scores

	Scale Reliability		Participant Scores	
	N of Items	Cronbach's α	Mean	(SD)
BIS-11	30	.81	60.91	(9.60)
Subscales				
Non-Planning Impulsiveness	11	.65	23.20	(4.37)
Motor Impulsiveness	11	.63	20.68	(3.99)
Attentional Impulsiveness	8	.73	17.03	(3.80)

Note. SD = Standard deviation.

Table 4

Mean Spearman's Rho and Pearson Correlations among Impulsiveness and Preference for Sequences for Each Category

			Pearson Correlation (r) Between r_s and			
	Spearman's Rho (r_s)		BIS-11		BIS-11 Non-planning	
Category	*M*	(SD)	*r*	*p*	*r*	*p*
Food	.20	.83	-.07	.36	-.01	.86
Noxious	.24	.79	.00	.95	-.01	.93
Exercise	.30	.78	.02	.81	.16	.03
School Work	.33	.78	-.04	.54	.01	.90
Leisure	.41	.75	-.06	.43	.00	.95
Mixed	-.04	.37	.00	.97	.11	.13

Note. Spearman's Rho (r_s) is the rank-order correlation between preference rank and position in the sequence. r = Pearson correlation. p = significance.

Study 2: Correspondence Between Sequence Preferences of Hypothetical and Real Outcomes

The purpose of Study 2 was to determine whether there was correspondence between college students' preference for sequences when the sequences involved hypothetical and real outcomes. Specifically, we aimed to determine whether time preference and the likelihood of saving the best for last differed when the sequences involved hypothetical or real school tasks, food, leisure activities, noxious stimuli, exercises, or a mix of all the categories. If there is correspondence, then one can be more confident when interpreting results of studies of time preference using hypothetical outcomes. If time preference differs when the outcomes are real versus hypothetical, then one will need to exercise caution when drawing conclusions on real human decisions from hypothetical studies of intended behavior.

Method

Participants, Setting, and Materials

Four men and four women between 18 and 24 years of age participated in Study 2. Participants were recruited from undergraduate psychology courses and invited to participate in the study via a psychology participant pool for extra credit and a chance to win a $25 Amazon gift card. Sessions were conducted in a laboratory space at the university, with a table, a computer, and all category-specific materials. In the sessions involving real outcomes, there were five laminated cards depicting numbers one through five for participants to use when ranking the items in the list. Materials necessary to complete the activities in each category were present.[v] In all sessions, there was a timer that the experimenter used to keep track of activity duration as well as pencils and paper data sheets for data collection.

Procedures[vi]

Overview. Participants were asked to come to the laboratory for two sessions. In the first session, which lasted approximately 30 min, participants had brief exposure to all the items that would be used in the rest of the study and completed the survey of hypothetical outcomes. In the second session, which lasted approximately 2 hr, participants ranked and generated sequences in which they would like to experience real outcomes and completed an exit survey.

Session 1. Participants sampled all items or activities for all categories, with up to 30 s access to each item. Duration of pre-exposure to an item or activity was never greater than the duration of the exposure during the actual test conditions. That is, if the exposure in the test condition was 5 s (e.g., cold-pressor task), then the pre-exposure was also limited to 5 s.

Following pre-exposure, participants completed the Qualtrics Sequence Survey from Study 1, on a laboratory computer. The list of food items was edited such that only the items that were pre-selected by the participant at the time of enrollment were included.

Session 2. In the second session, participants completed the leisure, noxious, exercise, food, and work categories in a randomly assigned order first, and ended with the mixed category. Because the mixed category consisted of the top-ranked item from each of the other categories, the mixed category had to be completed last. For each category, all items were displayed on the table and participants were shown each item briefly. Participants were then asked to rank their preference for each option, by placing the corresponding number card in front of each alternative (1 = highest preferred; 5 = least preferred). The experimenter then stated: “For the next [condition specific amount of time], you will have [specific time] to do each of these activities. Please tell me in what order you would like to do them.” The experimenter wrote down the sequence and showed it to the participant to confirm the desired sequence. If there were any errors in the recording of the preferred sequence, participants were given one opportunity to modify the sequence at this time. Once the preferred sequence was confirmed, the experimenter gave the materials for the first activity to the participant and started the timer for the appropriate duration

Leisure category. Leisure activities were presented, in the order selected, for 5 min each. When the 5 min elapsed with the first activity, the experimenter stated “Time is up. You now have 5 min to do [the next activity]” and removed the items for the first activity and presented the items for the activity scheduled second. The last two steps were repeated until the participant experienced all activities.

Noxious category. Prior to initiating the session, all items were displayed on the table. The experimenter stated: “For the next few min, you will experience several annoying or slightly uncomfortable stimuli. These are used as part of standard neurological testing in clinical and research settings, as well as to assess pain sensitivity. Other than discomfort, or slight pain, there is no real risk to your health or well-being in experiencing these. In case you do not remember them from last session, I’ll show you what they look like, so you know what to expect.” Participants were shown each item, one at a time, as the experimenter briefly described how the item was used. Participants were then asked to rank their preference for each option. The experimenter then stated:

"Over the next 2-5 min, you will experience five different events, each should take a few seconds, no more than 30 seconds. Please tell me in what order you would like to do them." Once the preferred sequence was confirmed, the experimenter proceeded with the first scheduled stimulus followed by the next in the sequence until all five had been presented. Participants could refuse to experience any of the stimuli presented, however, this never occurred.[vii] Except for the cold pressor task, which involved the entire hand and part of the forearm, all other stimulations were applied to the front of the wrist, 2.5 cm from the palm. This spot is also less likely to have sweat which could affect the sensory experience. Specific procedures for each sensory experience are described next.

Pin prick. The experimenter applied a light pin prick to the inside of the wrist one time for less than 1 s, applied from 5 cm away with moderate force with a single-use plastic pin made for use during neurological exams. The pin was designed not to break the skin.

Cold pressor task. The experimenter prompted participants to place a hand in a stainless-steel bucket with iced water (water temperature set to 3°, ± 1°C) one time for 5 s. The experimenter timed the 5 s using a stopwatch and prompted participants to remove the hand from the cold pressor when the timer expired.

Deep pressure. The experimenter applied the algometer (Wagner model FDX) to the participant's inner wrist and once the pressure display reached 1.8 kg, a 5-s timer was initiated. The experimenter removed the stimulus when the timer elapsed.

Repeated Von Frey. The experimenter applied the monofilament (60 g) against the participant's skin until the filament bended approximately halfway 30 times at 1 Hz, which is 30 times in 30 s for 1 s across applications in the same spot. The *repeated* Von

Frey is used as a test of increased pain perception to a repetitive stimulus (Barney, Hoch, Byiers, Dimian, & Symons, 2015). The experimenter stated “on” when applying the first touch and the data collector started a 30-s timer. The rest of the 30 touches were silently counted by the experimenter (i.e., 1-Mississippi, 2-Mississippi, …, 29-Mississippi, etc.) with each second being associated with a touch of the Von Frey monofilament to the skin. After 30 applications were complete, the experimenter said “off” to signal the end of the stimulus session, and the data collector stopped the timer. The data collector independently took data on the number of applications during the interval, to ensure integrity and calculate interobserver agreement (IOA).

Warm thermal probe. The experimenter touched the participant’s inner wrist with an electronic thermal heat probe, heated to 50°C (122°F) one time for 5 s. The experimenter stated "on" upon contact of the probe with the participant’s skin, timed the 5 s using a timer, and removed the stimulus when the timer elapsed.

Exercise category. Picture cards depicting the five different exercises were displayed on the table. The participant was shown each card and the experimenter reminded the participant how the exercise was performed by showing a brief video clip. Participants were then asked to rank their preference for each option. The experimenter then stated: “For the next 10 min, you will have 1 min to do each of these exercises, with a 1-min break in-between exercises. Please tell me in what order you would like to do them.” Once the preferred sequence was confirmed, the experimenter showed the picture card to the participant, stated “You have 1 min to do as many [exercises] as you can” and started the 1-min timer. When the 1 min elapsed, the experimenter stated “Time is up. You can take a 1-min break.” After the break, the experimenter stated, “You now have 1 min to do as many [the next exercise] as you can" and presented the items (e.g., picture card, floor mat) for the exercise scheduled second. The last two steps were repeated until the participant had experienced all activities.

Food category. The items that were pre-approved by the participant during enrollment in the study were displayed on the table. The participant was shown each one briefly, one at a time. Participants were then asked to rank their preference for each option. The experimenter then stated: "You will now sample each of these. Please tell me in what order you would like to sample them." Once the preferred sequence was confirmed, the experimenter administered a small sample of the first scheduled food to the participant. When the participant had consumed the food, the experimenter presented the next item. Participants could elect not to consume a food item; however, this never occurred.[viii]

Work category. All items were displayed on the table and the participant was shown each one briefly, one at a time. Participants were then asked to rank their preference for each option. The experimenter then stated: "For the next 25-30 min, you will have 5 min to work on each of these activities. Please tell me in what order you would like to do them." Once the preferred sequence was confirmed, the experimenter stated: "complete this as if it were a school assignment for which you would be graded. Do your best work," provided the materials for the first scheduled activity to the participant and started the 5-min timer. When the 5 min elapsed, the experimenter removed the materials for the current activity and said "Time is up. You now have 5 min to work on [the next activity]. Complete this as if it were a graded school assignment. Do your best work," and presented the items for the activity scheduled second. These steps continued until the participant had experienced all activities.[ix]

Mixed category. Outside of the laboratory, choices and daily schedules often involve a mix of activities of different categories. Therefore, the last category involved a mixed category, in which the top-ranked items from each of the preceding five categories were included. At the start of session, all items were displayed on the table and the participant was shown each one briefly, one at a time. Participants were then asked to rank their preference for each option. The experimenter then stated: "For the next few min, you will experience each of these options. Please tell me in what order you would like to experience them." Once the preferred sequence was confirmed, the experimenter administered the materials for the first scheduled item to the participant. If it was a leisure or work task, the experimenter started the 5-min timer. If it was exercise, the experimenter started a 1-min timer. When the participant had consumed the food, experienced the noxious stimulation, or when the timer elapsed, the experimenter said "Ok. You are all done with [activity]. You now get to [perform next activity]" and presented the items for the activity scheduled second. These steps continued until the participant had experienced all activities.[x]

Data Collection and Interobserver Agreement

For hypothetical outcomes, ranks for each item in the list and preferred sequence order were collected directly through the Qualtrics Sequence Survey. Pilot testing was conducted to ensure the Qualtrics data output matched participant responses.

For sessions involving real outcomes, two observers independently collected data during all sessions and used paper data sheets to record the ranks assigned to each item on the list as well as the sequence generated by the participant, and whether the participant requested changes to the sequence. For food items, data were collected on consumption or refusal of each item. For noxious experiences, data were collected on experimenter behavior (i.e., whether the experimenter implemented the stimuli as indicated), and participant behavior, specifically, completion of the experience or refusal. IOA for rank, sequence generated, food consumed, and experience, refusal, and integrity of noxious experiences was calculated by dividing the number of agreements by the number of agreements plus disagreements. There was 100% agreement across all the aforementioned measures. Engagement data were collected using a 10-s momentary time sampling for leisure and school work, and 5-s momentary time sampling for exercise. Specifically, the observation time was divided into 10-s (or 5-s) intervals, and observers recorded whether the participant was engaging with the activity at the end of the 10-s (or 5-s) interval (Becraft, Borrero, Davis, & Mendres-Smith, 2016). Because the exercise duration was just 1 min, the sampling intervals were reduced to allow for more observations. Engagement was defined as the participant holding the materials and manipulating them in a manner as it was intended (i.e., not just resting a hand on them), or the participant attempting to complete the activity, even if incorrectly. Interval-by-interval IOA for engagement with the outcome was calculated by adding the number of intervals with agreements and dividing it by the total number of intervals in the observation period (Becraft et al., 2016). Across participants, mean IOA for engagement in each category was 99.8% (range, 98% – 100%) for leisure, 99.6% (range, 96.7% – 100%) for exercise, 99.9% (range, 99.3% – 100%) for school work, and 99.7% (range, 98.6% – 100%) for the mixed category.

Results and Discussion

How do categorical features of items in an array influence saving the best for last and general sequence preference? For each sequence category, the researchers calculated the number of participants who chose the top-ranked option last or first in the sequence, as well as a Spearman rank order correlation between the relative value of the outcome, determined by the rank assigned by the participant, and the order in which it was selected. This correlation was intended to capture individual variations in sequence preferences. The resulting scores would range from −1 (perfectly improving, rank is inversely related to position in sequence) to 1 (perfectly worsening, rank is directly related to position in sequence) with scores closer to zero indicating a preference for interspersing higher and lower preferred outcomes. Results of these analyses are depicted in Table 5, separated by participants, and summarized for the entire sample. In general, participants were more likely to schedule the best first than last, except for real noxious outcomes, and hypothetical mixed sequences. For real leisure outcomes, an equal number of participants scheduled the best first or last. The greatest number of participants who saved the best for last (i.e., half of the sample) was observed with noxious stimuli when those were real. No participants saved the best for last when scheduling hypothetical exercises. Except for leisure and mixed categories, more participants saved the best for last when the outcomes were real rather than hypothetical. Given that all participants responded to the survey of hypothetical outcomes first and then the real outcomes, it is possible that the sequence of exposure can account, to some extent, for the fact that saving the best for last was never less likely with real outcomes than it was with hypothetical outcomes.

Table 5

Saving the Best for Last, Scheduling the Best First and General Sequence Preference by Category and Type of Outcome for Study 2

	Participant																Summary	
	#1		#2		#3		#4		#5		#6		#7		#8		*f* (%)	
	HO	RO	HO	RO	HO	RO	HO	RO	HO	RO	HO	RO	HO	RO	HO	RO	HO	RO
Noxious																		
Best Last	0	0	1	1	0	0	0	1	1	1	0	0	0	1	0	0	2 (25)	4 (50)
Best First	0	0	0	0	1	1	1	0	0	0	1	0	1	0	1	1	5 (63)	2 (25)
r_s	-.6	-.2	-1.0	-1.0	.7	1.0	.9	-1.0	-.1	-.3	.7	-.2	.9	-1.0	1.0	1.0		
Leisure																		
Best Last	1	0	1	0	0	0	1	1	0	0	0	1	0	1	0	0	3 (38)	3 (38)
Best First	0	0	0	0	1	1	0	0	1	1	1	0	1	0	1	1	5 (63)	3 (38)
r_s	-.4	.3	-.9	.1	.9	1.0	-1.0	-1.0	1.0	1.0	.9	-.9	.8	-1.0	1.0	1.0		
School Work																		
Best Last	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	1 (13)	2 (25)
Best First	0	0	0	1	1	1	1	0	1	1	1	1	1	0	1	1	6 (75)	5 (63)
r_s	-.8	-.4	-.4	.6	.9	1.0	1.0	-1.0	.3	1.0	1.0	1.0	1.0	-1.0	1.0	1.0		
Food																		
Best Last	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1 (13)	2 (25)
Best First	1	1	0	1	1	0	1	0	1	1	1	1	1	1	1	0	7 (88)	5 (63)
r_s	.1	.8	-.6	.1	1.0	-1.0	1.0	-1.0	.9	1.0	1.0	1.0	1.0	1.0	1.0	-.7		
Exercise																		
Best Last	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0 (0)	1 (13)
Best First	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	4 (50)	5 (63)
r_s	.3	.6	1.0	1.0	1.0	1.0	-.3	-1.0	.7	1.0	.9	.8	.0	1.0	1.0	1.0		
Mixed																		
Best Last	1	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	2 (25)	2 (25)
Best First	0	0	0	1	0	1	0	0	0	1	0	0	0	0	1	0	1 (13)	3 (38)
r_s	-.7	-.3	.3	.1	-.1	1.0	-.6	-.9	.1	1.0	.5	-.1	.4	-1.0	.4	.9		

Note. N = 8. HO = Hypothetical Outcomes; RO = Real Outcomes; *f* = frequency of participants; Best Last = whether the participant saved the best for last; Best First = whether the participant scheduled the best option first in the sequence; r_s = Spearman rank-order correlation between preference rank and position in the sequence.

What degree of correspondence is observed in responding to hypothetical or real outcomes? For each participant, the researchers calculated a Pearson correlation between participant's ranks of hypothetical outcomes and their ranks of the real outcomes. Similarly, the researchers calculated a Pearson correlation between the sequences generated for hypothetical outcomes and those of the real outcomes. The results of the correspondence analysis for ranks and sequences are depicted graphically in Figure 1 as the percentage of responses that had a strong positive correlation (defined as $r \geq .6$), a weak positive correlation ($.5 \leq r \geq .2$), no correlation ($.1 \leq r \geq -.1$), a weak negative correlation ($-.2 \leq r \geq -.5$), or a strong negative correlation ($r \leq -.6$). Overall, there was very strong correspondence in the ranks assigned to the hypothetical and real outcomes, suggesting that people's relative preference for outcomes is consistent.

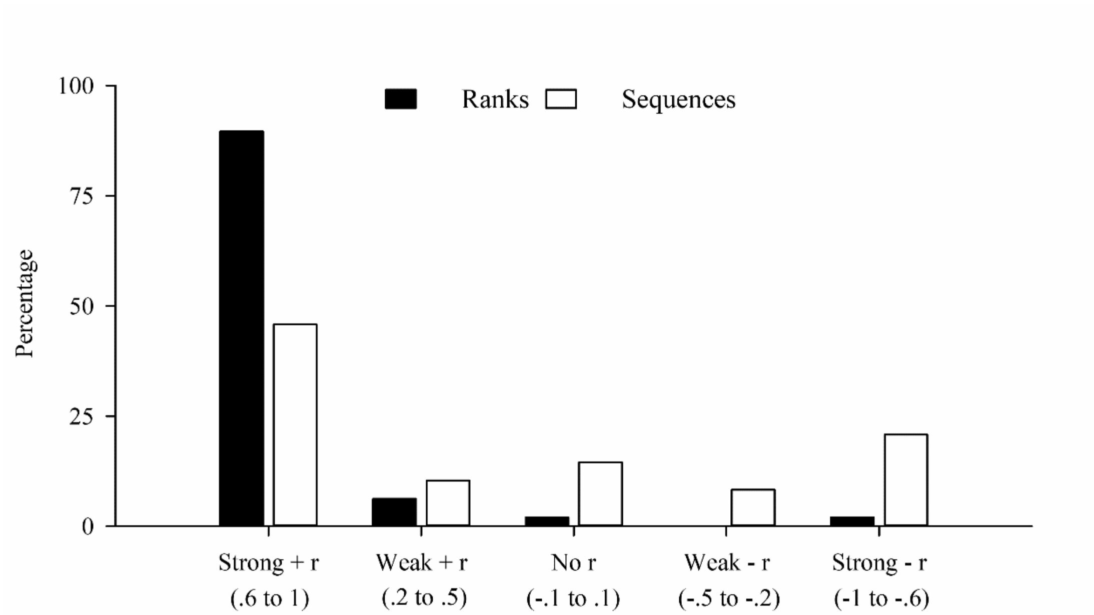

Figure 1. Correspondence between hypothetical survey and real outcomes. There was more variability in the correlations between the sequences generated, which suggests that although rank may be constant across type of outcome and time of response, sequence

preferences may be more variable. It is important to note that in the current study, participants responded to the surveys just once. Furthermore, during the exit survey, one of the participants stated that she would have scheduled activities differently were she to do it again. Future studies could conduct repeated measures and compare consistency across measures and correspondence between responses to hypothetical and real outcomes given repeated opportunities to respond.

Study 3: Outcome Category as a Predictor of Preschooler's Time Preference

Prior studies have found that children's preference for delayed rewards increases with age, with kindergarten children being the least likely to select a larger later reward over a smaller sooner reward, compared to elementary school children (e.g., Mischel & Metzner, 1962). In studies of delayed gratification, in which children can obtain a less preferred reward immediately or continue waiting indefinitely for a more preferred reward, considerable variability is often observed in the amount of time preschool children voluntarily delay gratification (e.g., Mischel, Shoda, & Rodriguez, 1989).

Given the lack of studies on children's preference for sequences, it is unclear whether negative time preference, or saving the best for last, emerges along a developmental path that parallels the development of self-control (e.g., Mischel & Mischel, 1983). Although delay of gratification may be less likely with preschoolers, it is possible that framing the choice as a sequence may function as a strategy that promotes saving the best for last. The purpose of this study was to determine whether typically-developing preschool children also choose to save the best for last and whether the likelihood of exhibiting negative time preference differs when the sequences involved school tasks, food, exercise, books, leisure activities, or a mixed category.

Method

Participants, Setting, and Materials

Four typically-developing preschoolers participated in Study 3. Participants were recruited from a preschool in the mid-Atlantic region. Cole and David were 4.5 years old at the time of the study and were in the four-year-old classroom. Andrea and Carla were 5 years old at the time of the study and were recruited from the five-year-old classroom. Sessions were conducted in a quiet room in the preschool (i.e., the staff lounge) furnished with two chairs and a table. Session materials included a timer and the materials relevant to specific categories. Additionally, there were picture cards depicting each of the activities or items in the different categories that participants used to create a visual schedule when generating their preferred sequence order. Items for the different categories were selected in consultation with the preschool director, preschool teachers, and parents. The work category included tracing or coloring sheets, a lacing task, a shape identification task, a counting task, and a letter identification task. The food category included items that were pre-approved by the participant's parent or guardian and included blueberries, fruit snacks, veggie straws, carrots, yogurt, chips, and crackers. The leisure category included a bin with uncooked pasta for children to play with, Lego Duplos®, costumes for pretend play, Incredibles figurines, and Play-Doh®. The story-time category included five different children's story books. The exercise category included an exercise mat and a hula hoop.

Procedures

General procedures. During the informed consent process, parents were asked to disclose if their child had any food allergies. Parents were given a list of the foods that could be used during session so they could indicate their acceptance of the food choices and write down any additional recommendations.

Participants were asked to take part in seven sessions, lasting 5 to 30 min each. The first session was used to build rapport with the participant, and to give the participant exposure to the different items and activities that would be used throughout the study. During that first meeting, a brief delay sensitivity assessment was conducted. Specifically using a Hershey's kiss, which had been identified by the caregivers as highly preferred, the experimenter told the participant: "We have a chocolate for you. Do you want this chocolate now or later, in about 3 minutes?" If the participant selected "now," the item was delivered immediately, and the assessment ended. If the participant selected "later" the experimenter started a 3-min timer and said "OK, you chose in 3 min. You can have your chocolate when the timer beeps." When the 3-min timer elapsed, the experimenter delivered the chocolate to the participant, and then repeated the procedure two more times, for a total of three trials.

Each of the remaining six sessions assessed a different category. Prior to starting, the order of the categories was randomized, but the mixed category was always last. In each session participants (a) generated a preference rank hierarchy in a paired-choice format, (b) generated a preferred sequence with a visual schedule, and (c) experienced all the activities in the scheduled order.

Preference ranking. At the beginning of session, the participant was given brief (approximately 5 s) exposure to each activity while the experimenter paired it with the corresponding picture card. Following pre-exposure, the experimenter stated: "I want to get to know you a little better and see what you like. I will show you a few things, and I want you to tell me which one you like the most. We will get to do these in a little while." Participants then completed a modified paired-stimulus preference assessment (PSPA, Fisher et al., 1992), explained next, with the picture icons that represented the actual materials. For each possible pairwise combination, participants were asked: "which of these do you like the most [item x] or [item y]?" Once identified, the experimenter said "Okay, so you like [item] more than [other item]" then moved on to the next pair. Following selection, participants could hold the picture card briefly, but did not get access to the items during the modified PSPA. Although not providing immediate access to the selected item is atypical in a PSPA, it has been successfully carried out in the past with children with and without high-functioning autism (Goldberg et al., 2017). Restricting access during the PSPA was intended to minimize satiation with the items, which could later affect how the participant generated a preferred sequence. Furthermore, only one round of the PSPA was conducted, for a total of 10 trials, making the entire assessment relatively brief.

Preferred sequence. Once a hierarchy was identified, participants were asked to generate a sequence by placing the picture icons on a visual schedule, in the order in which they would like to experience them. The experimenter stated: "For the rest of our time together, we get to [do, read, or eat] all these. Let's create our own schedule. What do you want to do first? Put the picture card at the top of the schedule board…What do you want to do second?" and so on. Once all the cards had been placed on the schedule, the experimenter verbally described the schedule order, "Okay, you said you want to [perform activity] first, then [next activity]" and so on. At this point, the participant could agree to the schedule, or choose to change it one time. The participants then completed the activities for the corresponding category, as described below. Participants were encouraged to engage in an activity throughout the duration. If the participant stopped engaging in the activity, the experimenter continued to provide verbal prompts or statements of encouragement until the timer elapsed. At the end of each session, participants were given a small sticker. Following completion of the study, participants were given a small thank-you gift (e.g., a book, bubbles) that was up to $10 in value.

Leisure and work categories. The experimenter pointed to the first icon on the schedule and said "Okay. We will be starting with [activity]." The experimenter then gave the materials for the first activity to the participant and started the 2- to 4-min timer. The duration of access to the activities was 4-min for all participants, except for David's work category. Academic activities were reportedly very aversive to David, therefore, to minimize the chances of making the whole experience very aversive, the duration of the academic activities was restricted to 2 min for him. When the timer elapsed, the experimenter said "Okay, we are all done with [activity]. You can place the picture icon in the 'all done' pocket" and removed the materials for the first activity. The experimenter then said "Next, we have [next activity]" and presented the items for the activity that was placed second in the sequence. These steps continued until the participant had experienced all activities.

Story-time category. The experimenter pointed to the first icon on the schedule and said "Okay. We will be starting with [book]," and read the book with the participant. When the book was over, the experimenter said "Okay, we are all done with [book]. You can place the picture icon in the 'all done' pocket. Next, we have [next book]," and started reading the book that was scheduled second in the sequence. These steps continued until the participant had read all books.

Exercise category. The experimenter pointed to the first icon on the schedule and said "Okay. We will be starting with [exercise]," started a 30-s timer and engaged in the exercise along with the participant. When the timer elapsed, the experimenter said "Okay, we are all done with [exercise]. You can place the picture icon in the 'all done' pocket. Next, we have [next exercise]," and prompted the participant to do the exercise that was scheduled second in the sequence. These steps continued until the participant had performed all exercises.

Food category. The experimenter pointed to the first icon on the schedule and said "Okay. You get to eat [food]," and gave a small piece of the first food to the participant. When the participant had consumed the food, the experimenter presented the next item. These steps continued until the participant had sampled all foods.

Mixed category. The top-ranked items from each of the preceding categories were included. The experimenter pointed to the first icon on the schedule and said "Okay. We will be starting with [item]." The experimenter gave the materials for the item to the participant. If it was a leisure or work task, the experimenter started 4-min timer (or a 2-min timer for David's work task). If it was an exercise, the experimenter started the 30-s timer. When the participant had consumed the food, the experimenter has finished reading the book, or when the timer elapsed, the experimenter said "Okay. We are all done with [item]." The experimenter removed the materials and prompted the participant using the following phrase "You can place the picture icon in the 'all done' pocket. Next we have [activity]." The experimenter then presented the items for the second scheduled activity. These steps continued until the participant had experienced all activities.

Data Collection and Interobserver Agreement

Observers used paper data sheets to record selection during the PSPA and the sequence generated by the participant. Observers recorded how much time it took participants to generate a sequence (from initial prompt until the last picture card was placed on the schedule), and whether the participant choose to change it when give the option (prior to experiencing any of the activities). For food items, data were collected on consumption or refusal of each item. Engagement data were collected using a 10-s momentary time sampling for leisure, story time, and school work, and 5-s momentary time sampling for exercise. Specifically, the observation time was divided into 10-s (or 5-s) intervals, and observers recorded whether the participant was engaging with the activity at the end of the 10-s (or 5-s) interval (Becraft et al., 2016). Because the exercise duration was just 30 s, the sampling intervals were reduced to allow for more observations. Engagement was defined as the participant holding the materials and manipulating them in a matter as it was intended (i.e., not just resting a hand on them or touching them to throw across the room; Keen & Pennell, 2010; Peters & Thompson, 2013), or the participant attempting to complete the activity, even if incorrectly. For story time, engagement was scored when the participant was facing the book or talking about the story. A second observer independently collected data during 100% of sessions across phases and participants. IOA for the PSPA, sequence generated, and food consumed was calculated by dividing the number of agreements by the number of agreements plus disagreements. There was 100% agreement on selection in all PSPA sessions for all participants. There was also 100% agreement in all sessions and for all participants regarding the sequence generated, whether the participant requested to change the sequence when given the option, and food consumption. Similarly, interval-by-interval IOA for engagement with the outcome was calculated by adding the number of intervals with agreements and dividing it by the total number of intervals in the observation period (Becraft et al., 2016). Mean IOA for engagement included 95.5% (range, 90.0% – 100%) for Cole, 99.5% (range, 98.7% – 100%) for Andrea, 89.3% (range, 69.2% – 100%) for David, and 100% for Carla.[xi]

Results and Discussion

During the delayed sensitivity assessment, Cole, David, and Carla selected the immediate option on the first trial. Andrea selected the delayed option (i.e., the same Hershey Kiss following a 3-min delay) in the first trial, while clarifying "because my mommy said I have to wait." Following the delay, she was given access to the Hershey Kiss. In the two subsequent trials, she selected the immediate option both times. Results of the delay sensitivity assessment suggest that our participants were in fact aware of the delays, and that, they exhibited positive time preference in this scenario, preferring immediate over delayed consumption. Results of the sequence preference evaluation are depicted in Table 6 and will be discussed next in the context of two specific questions.

How do categorical features of items in the array influence saving the best for last? This question was answered by specifically looking at whether participants saved the best option for last in their sequence for each category (see "STBFL" rows in Table 6). When the sequence involved academic-type activities, two of the participants (i.e., Cole and Carla) scheduled the best for last. When the sequence involved exercise, story time, food, or a mix of all the categories, Carla was the only one of the four participants who saved the best for last. None of the participants saved the best for last when the sequence consisted exclusively of leisure items. Conversely, in almost all categories (food category excluded), half of the participants chose to schedule their favorite activity first (see "Best First" rows in Table 6). Overall, in 11 of the 24 sequences, the best activity was scheduled first, whereas the best was saved for last in 6 of the 24 sequences.

How do the categorical features of items in an array influence sequence preference? For each sequence category we calculated a Spearman rank order correlation between the relative value of the outcome, determined by the rank assigned by the participant, and the order in which it was selected. This correlation is intended to capture individual variations in sequence preferences. The resulting scores range from −1 (perfectly improving, rank is inversely related to position in sequence) to 1 (perfectly worsening, rank is directly related to position in sequence) with scores closer to zero indicating a preference for interspersing higher and lower preferred outcomes. The correlation coefficients are shown in the third row for each sequence category in Table 6. Results for individual participants are depicted in their respective columns, with the results for the entire sample depicted in the summary section on the right. Although in 5 of the 6 categories at least one participant chose to schedule the best option last, none of the participants generated a perfectly improving sequence. Preference for generally improving and worsening sequences varied across participants and most categories.

Table 6

Saving the Best for Last, Scheduling the Best First and General Sequence Preference by Category for Study 3 (Preschoolers)

	Participant				Summary			
	Cole	Andrea	David	Carla	*f*	(%)	*M*	(SD)
Academics								
Best Last	1	0	0	1	2	(50)		
Best First	0	1	1	0	2	(50)		
r_s	-.30	.10	.40	-.60			-.10	(.44)
Story Time								
Best Last	0	0	0	1	1	(25)		
Best First	1	0	1	0	2	(50)		
r_s	.43	-.45	.10	-.60			-.13	(.48)
Food								
Best Last	0	0	0	1	1	(25)		
Best First	1	0	0	0	1	(25)		
r_s	.70	-.10	.40	-.90			.03	(.70)
Exercise								
Best Last	0	0	0	1	1	(25)		
Best First	1	0	1	0	2	(50)		
r_s	.30	.80	.70	-.89			.23	(.78)
Mixed								
Best Last	0	0	0	1	1	(25)		
Best First	1	0	1	0	2	(50)		
r_s	.70	.80	.50	.00			.50	(.36)
Leisure								
Best Last	0	0	0	0	0	(0)		
Best First	1	0	1	0	2	(50)		
r_s	1.00	.50	.67	.70			.72	(.21)

Note. N = 4. *f* = frequency of participants; Best Last = whether the participant saved the best for last; Best First = whether the participant scheduled the best option first in the sequence; r_s = Spearman rank-order correlation between preference rank and position in the sequence.

When the sequences involved academics, story time, food, and exercise, participants for the most part chose to intersperse more preferred activities with less preferred activities. Furthermore, in the mixed category, 3 of the 4 participants generated generally worsening sequences, and one participant -Carla- generated a mixed sequence with the first four scheduled activities going from more to less preferred (i.e., perfectly worsening), but saving the best activity until the end, or fifth in the sequence (hence, r_s = 0). In the leisure category, however, all participants generated generally worsening sequences ($r_s > 0$). Therefore, interspersing more- and less-preferred activities seems most likely with academics, story time, food, and exercise. Preference for interspersing, but generally worsening sequences seems likely with mixed sequences. Preference for worsening sequences or experiencing the most preferred sooner seems most likely with leisure items.

General Discussion

We investigated how young adults and preschool children choose to schedule outcomes of different categories. In the current studies, roughly a quarter of participants chose to save the best for last. This is quite different from the results of our previous study on sequence preferences in which 64.87% of participants saved the best for last when arranging a sequence of five hypothetical outcomes (Castillo et al., 2020). One possible explanation for this discrepancy is the procedural variations in the studies. Specifically, in the present Study 1 and Study 2, participants ranked and scheduled textual descriptions of different items, activities, or stimuli. Conversely, in the study by Castillo et al. (2020), participants were not provided with specific items, activities, or stimuli to rank, but were prompted to think of brief activities they liked to different degrees. Participants then generated a sequence using bars of different heights to represent relative quantitative or qualitative differences in options. Prior research has suggested that when choosing sequences of unavoidable outcomes, people tend to prefer improving sequences (Loewenstein & Prelec, 1991), and that this may be particularly true when aversive outcomes are involved (Story et al., 2013). Results of the current studies suggest that this preference for sequences, or saving the best for last, may not be as robust as previously indicated. Furthermore, these results provide support for prior conclusions that different methods may occasion different response patterns (Andrade & Hackenberg, 2012). Using graphical representations of sequences (e.g., bars of different heights), may encourage participants to apply a more superficial level of analysis in which preference for improvement is more likely (Frederick & Loewenstein, 2008). Nevertheless, despite a significantly smaller percentage of participants who saved the best for last in the current Study 1, the percentage of participants who saved the best for last was highest in the hypothetical food and noxious categories, and in Study 2, half of the participants saved the best for last when scheduling *real* noxious stimuli. Correspondence between responses to hypothetical or real outcomes was generally strong for preference ranks, but more variable for sequences generated. It makes sense that relative preference for outcomes remains constant in the short term, even if the scheduling preference changes from moment-to-moment.

To our knowledge, preference for sequences had not been evaluated with preschoolers before. Although there is reason to believe, as was supported by our findings, that preschoolers will want their preferred items and activities sooner, 1 of 4 participants chose to save the best for last in all categories, except for the leisure. The authors recognize that given the size of the sample one cannot make generalizing conclusions. Nevertheless, the “save the best for last” phenomenon *was* documented despite having run this with only four children.

In Study 2, we inferred preference based on the sequences participants generated prospectively. That is, following exposure to the different outcomes, participants generated a sequence, and had to commit to the order at the onset of the session. Because we did not provide a list of predetermined sequences to choose from, or experience, one cannot say anything about the relative preference for one arrangement, or sequence order, over another. Moreover, we did not ask participants to retrospectively evaluate the sequences experienced because in the current study there was no relative comparison (i.e., participants only experienced the sequence they generated). Nevertheless, research on retrospective evaluations of experiences extended across time suggests that evaluations of experiences are heavily influenced by the average of the peak (e.g., worst) and end (i.e., final) moments of the experience (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993).[xii] Therefore, findings from prospective or retrospective evaluations of sequences may indicate a general preference for improvement. Consequently, even if the overall pain experienced would not differ, individuals undergoing painful medical treatments may prefer treatments that are ordered in a way such that the pain diminishes over time, rather than gradually increases, and their memory of the treatment may be less aversive if relief from the pain is gradual rather than abrupt (Kahneman et al., 1993). In Study 2, half of the participants saved the best for last when scheduling noxious stimuli, and at least one participant stated in the exit survey that if she were to do it again, she would have scheduled it from least to most preferred. The fact that young adults exhibit negative time preference with noxious outcomes may be worth noting when scheduling medical interventions within a single session. This consideration is important because it is possible that retrospective subjective evaluations of a medical appointment or intervention could influence a person's morale or even compliance with follow-up appointments and future treatment recommendations. This could all subsequently affect medical outcomes (Redelmeier & Kahneman, 1996). Of course, further extensions are necessary to determine if these things would occur.

In the current studies participants ranked and generated sequences just one time per category or type of outcome. It is possible however, that results would have differed with repeated measures. The authors believe there are many possible investigations that could be conducted regarding preference for sequences. For example, future studies could evaluate (a) response patterns when participants can generate sequences multiple times, (b) retrospective evaluations of sequences, when those are generated by the participants or by the experimenter, and (c) whether experiencing sequences in a preferred order is also associated with changes in performance during the different activities.

Finally, the current series of studies investigated whether college students and typically developing preschoolers save the best for last when scheduling sequences of categorically-different outcomes. Saving the best for last was observed across both age groups, even if to a smaller degree than has been reported in prior research. Furthermore, outcome category does seem to influence the percentage of participants who saved the best for last. One question that has not been answered yet is whether individuals with intellectual and developmental disabilities will also save the best for last when presented with an array of items with which to engage. The opportunity to choose the order of exposure is embedded in preference assessment procedures commonly used for individuals with intellectual and developmental disabilities (e.g., the Multiple Stimulus Without Replacement, DeLeon & Iwata, 1996, and the free operant preference assessment, Roane, Vollmer, Ringdahl, & Marcus, 1998). The item selected first in these assessments is often incorporated into behavioral programming to reduce problem behavior, increase appropriate behavior, or both. However, if individuals with intellectual and developmental disabilities are saving the best for last, then behavioral interventions may not involve the most preferred outcomes. At present, we do not know whether individuals with intellectual and developmental disabilities save the best for last when selecting items in a Multiple Stimulus Without Replacement arrangement (Deleon & Iwata, 1996). Nevertheless, an answer to this question could have meaningful implications for the way in which we conduct or interpret the results of preference assessments.

References

Admon, R., & Pizzagalli, D. A. (2015). Dysfunctional reward processing in depression. *Current Opinion in Psychology, 4*, 114-118. doi: 10.1016/j.copsyc.2014.12.011

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82*, 463-496. doi:10.1037/h0076860

Andrade, L. F., & Hackenberg, T. D. (2012). Saving the best for last? A cross-species analysis of choices between reinforcer sequences. *Journal of the Experimental Analysis of Behavior, 98*, 45-64. doi: 10.1901/jeab.2012.98-45

Ariely, D., & Carmon, Z. (2000). Gestalt characteristics of experiences: The defining features of summarized events. *Journal of Behavioral Decision Making, 13*, 191-201. doi: 10.1002/(SICI)1099-0771(200004/06)13:2<191::AID-BDM330>3.0.CO;2-A

Ariely, D., & Zauberman, G. (2000). On the making of an experience: The effects of breaking and combining experiences on their overall evaluation. *Journal of Behavioral Decision Making, 13*, 219-232. doi: 10.1002/(SICI)1099-0771(200004/06)13:2<219::AID-BDM331>3.0.CO;2-P

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

Barney, C. C., Hoch, J., Byiers, B., Dimian, A., & Symons, F. J. (2015). A case-controlled investigation of pain experience and sensory function in Neuronal Ceroid Lipofuscinosis. *Clinical Journal of Pain, 31*, 998-1003. doi: 10.1097/AJP.0000000000000192.

Barney, C. C., Tervo, R., Wilcox, G. L., & Symons, F. J. (2017). A case-controlled investigation of tactile reactivity in young children with and without global developmental delay. *American Journal of Intellectual and Developmental Disabilities*, *122*, 409-421. doi: 10.1352/1944-7558-122.5.409

Barratt, E. S. (1959). Anxiety and impulsiveness related to psychomotor efficiency. *Perceptual and Motor Skills, 9*, 191-198.

Becraft, J. L., Borrero, J. C., Davis, B. J., & Mendres-Smith, A. E. (2016). Assessment of a rotating time sampling procedure: Implications for interobserver agreement and response measurement. *Education and Treatment of Children, 39*, 1-20.

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

Bilaszek, W., Ostaszewski, P., Green, L., & Myerson, J. (2019). On four types of devaluation of outcomes due to their costs: delay, probability, effort, and social discounting. *The Psychological Record.* Advance online publication: doi: 10.1007/s40732-019-00340-x

Bryant, F. B., Chadwick, E. D., & Kluwe, K. (2011). Understanding the processes that regulate positive emotional experience: Unsolved problems and future directions for theory and research on savoring. *International Journal of Wellbeing*, *1,* 107-126. doi: 10.5502/ijw.v1i1.18

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology*, 67, 319-333. doi: 10.1037/0022-3514.67.2.319

Castillo, M. I, Sun, S., Frank-Crawford, M. A., & Borrero, J. C. (2020). *Save the best for last I: Young adults demonstrate negative time preference - A replication and extension.* Manuscript in preparation.

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

Cook, J. O., & Barnes, L. W., Jr. (1964). Choice of delay of inevitable shock. *Journal of Abnormal and Social Psychology*, *68*, 669-672. doi:10.1037/h0047127

De Wit, H., Flory, J. D., Acheson, A., McCloskey, M. & Manuck, S. B. (2007). IQ and nonplanning impulsivity are independently associated with delay-discounting in middle-aged adults. *Personality and Individual Differences*, *42*, 111-121. doi: 10.1016/j.paid.2006.06.026

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29*, 519-533. doi: 10.1901/jaba.1996.29-519

Drolet, A., Lau-Gesk, L., & Scott, C. (2011). The influence of aging on preferences for sequences of mixed affective events. *Journal of Behavioral Decision Making, 24*, 293-314. doi:10.1002/bdm.695

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality, 45*, 259–268. doi: 10.1016/j.jrp.2011.02.004

Farber, M. L. (1953). Time-perspective and feeling-tone: A study in the perception of the days. *The Journal of Psychology, 35*, 253-257. doi:10.1080/00223980.1953.9712859

FeldmanHall, O., Dalgleish, T., Thompson, R., Evans, D., Schweizer, S., & Mobbs, D. (2012). Differential neural circuitry and self-interest in real vs hypothetical moral decisions. *Social Cognitive and Affective Neuroscience, 7,* 743-751. doi: 10.1093/scan/nss069

Fisher, W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. *Journal of Applied Behavior Analysis, 25,* 491-498. doi:10.1901/jaba.1992.25-491

Frederick, S., & Loewenstein, G. (2008). Conflicting motives in evaluations of sequences. *Journal of Risk and Uncertainty, 37,* 221-235. doi:10.1007/s11166-008-9051-z

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature, 40,* 351-401. doi:10.1257/002205102320161311

Friedel, J. E., DeHart, W. B., Madden, G. J., & Odum, A. L. (2014). Impulsivity and cigarette smoking: Discounting of monetary and consumable outcomes in current and non-smokers. *Psychopharmacology, 231,* 4517-4526. doi: 10.1007/s00213-014-3597-z

Goldberg, M. C., Allman, M. J., Hagopian, L. P., Triggs, M. M., Frank-Crawford, M. A., Mostofsky, S. H., Denckla, M. B., & DeLeon, I. G. (2017). Examining the reinforcing value of stimuli within social and non-scoail contexts in children with and without high-functioning autism. *Autism, 21,* 881-895. doi: 10.1177/13623613166550

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

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science, 4*, 401-405. doi:10.1111/j.1467-9280.1993.tb00589.x

Kasl, S. V., Gore, S., & Cobb, S. (1975). The experience of losing a job: Reported changes in health, symptoms and illness behavior. *Psychosomatic Medicine, 37*, 106-122. doi:10.1097/00006842-197503000-00002

Kang, M. J., Rangel, A., Camus, M., & Camerer, C. F. (2011). Hypothetical and real choice differentially activate common valuation areas. *The Journal of Neuroscience, 31*, 461-468. doi: 10.1523/JNEUROSCI.1583-10.201

Keen, D., & Pennell, D. (2010). Evaluating an engagement-based preference assessment for children with Autism. *Research in Autism Spectrum Disorders, 4*, 645-652. doi: 10.1016/j.rasd.2009.12.010

Kumar, A., & Gilovich, T. (2016). To do or to have, now or later? The preferred consumption profiles of material and experiential purchases. *Journal of Consumer Psychology, 26*, 169-178. doi: 10.1016/j.jcps.2015.06.013

Kumar, A., Killingsworth, M. A., & Gilovich, T. (2014). Waiting for Merlot: Anticipatory consumption of experiential and material purchases. *Psychological Science, 25*, 1924-1931. doi:10.1177/0956797614546556

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*, 173-187. doi: 10.1016/j.beproc.2005.02.003

Locey, M. L., Jones, B. A., & Rachlin, H. (2011). Real and hypothetical rewards in self-control and social discounting. *Judgment and Decision Making, 6,* 552-564. Doi:

Löckenhoff, C. E., Rutt, J. L., Samanez-Larkin, G. R., O'Donoghue, T., & Reyna, V. F. (2019). Preferences for temporal sequences of real outcomes differ across domains but do not vary by age. *Journals of Gerontology: Psychological Sciences, 74*, 430-439. doi:10.1093/geronb/gbx094

Loewenstein, G. (1987). Anticipation and the valuation of delayed consumption. *The Economic Journal, 97*, 666-684. doi:10.2307/2232929

Loewenstein, G., & Prelec, D. (1991). Negative time preference. *The American Economic Review, 81*, 347-352.

Loewenstein, G. F., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review, 100*, 91-108.

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

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

Magen, E., Dweck, C. S., & Gross, J. J. (2008). The hidden-zero effect: Representing a single choice as an extended sequence reduces impulsive choice. *Psychological Science, 19,* 648-649. doi:10.1111/j.1467-9280.2008.02137.x

Marshall, L. L., & Kidd, R. F. (1981). Good news or bad news first? *Social Behavior and Personality, 9*, 223-226. doi:10.2224/sbp.1981.9.2.223

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*, 276-282. doi: 10.11613/BM.2012.031

Mischel, W., & Metzner, R. (1962). Preference for delayed reward as a function of age, intelligence, and length of delay interval. *Journal of Abnormal and Social Psychology, 64*, 425-431. doi: 10.1037/h0045046

Mischel, H. N., & Mischel, W. (1983). The development of children's knowledge of self-control strategies. *Child Development, 54*, 603-619. Doi: 10.2307/1130047

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science, 244*, 933-938. doi: 10.1126/science.2658056

Mitchell, L. A., MacDonald, R. A. R., & Brodie, E. E. (2004). Temperature and the cold pressor test. *The Journal of Pain, 5*, 233-238. doi: 10.1016/j.jpain.2004.03.004

Nunnally, J., & Bernstein, L. (1994). *Psychometric theory*. New York: McGraw-Hill Higher, Inc.

Odum, A. L. (2011). Delay discounting: I'm a k, you're a k. *Journal of the Experimental Analysis of Behavior, 96*, 427-439. doi: 10.1901/jeab.2011.96-423

Olson, M., & Bailey, M. (1981). Positive time preference. *Journal of Political Economy, 89,* 1-25. Retrieved from http://www.jstor.org.proxy-bc.researchport.umd.edu/stable/1837349

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology, 51*, 768-774.

Peters, L. C., & Thompson, R. H. (2013). Some indirect effects of positive practice overcorrection. *Journal of Applied Behavior Analysis, 46*, 613-625. doi: 10.1002/jaba.63

Qualtrics® XM (2018) [Online survey platform]. Provo, UT: Qualtrics.

Quoidbach, J., Wood, A. M., & Hansenne, M. (2009). Back to the future: The effect of daily practice of mental time travel into the future on happiness and anxiety. *The Journal of Positive Psychology, 4*, 349-355. doi: 10.1080/17439760902992365

Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain, 66*, 3-8. doi:10.1016/0304-3959(96)02994-6

Roane, H. S., Vollmer, T. R., Ringdahl, J. E., & Marcus, B. A. (1998). Evaluation of a brief stimulus preference assessment. *Journal of Applied Behavior Analysis, 31*, 605-620. doi: 10.1901/jaba.1998.31-605

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L. Anderson, N. E., & Patton, J. H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and Individual Differences, 47*, 385-395. doi: 10.1016/j.paid.2009.04.008

Story, G. W., Vlaev, I., Seymour, B., Winston, J. S., Darzi, A., & Dolan, R. J. (2013). Dread and the disvalue of future pain. *PLoS Computational Biology, 9*, 1-18. doi: 10.1371/journal.pcbi.1003335

Suddendorf, T. & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, *30*, 299-351. doi: 10.1017/S0140525X07001975

Tulving, E. (1984). Précis of elements of episodic memory. *Behavioral and Brain Sciences, 7,* 223-268. doi: 10.1017/S0140525X0004440X

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211,* 453-458. doi:10.1126/science.7455683

Vlaev, I. (2012). How different are real and hypothetical decisions? Overestimation, contrast and assimilation in social interaction. *Journal of Economic Psychology, 33,* 963-972. doi: 10.1016/j.joep.2012.05.005

von Baeyer, C. L., Piira, T., Chambers, C. T., Trapanotto, M., & Zelter, L. K. (2005). Guidelines for the cold pressor task as an experimental pain stimulus for use with children. *The Journal of Pain, 6,* 218-227. doi: 10.1016/j.jpain.2005.01.349

Yarnitsky, D. & Pud, D. (2004). Quantitative sensory testing. In C. D. Binnie, R. Cooper, F. Mauguière, J. W. Osselton, P. F. Prior, & B. M. Tedman (Eds.), *Clinical Neurophysiology (Revised and Enlarged Edition): Vol. 1. EMG, Nerve Conduction and Evoked Potentials* (pp. 309-336). Amsterdam: Elsevier.

Xu, S., Pan, Y., Qu, Z., Fang, Z., Yang., Z., Yang, F., Wang, F., & Rao, H. (2018). Differential effects of real versus hypothetical monetary reward magnitude on risk-taking behavior and brain activity. *Science Reports, 8,* 3712. doi: 10.1038/s41598-018-21820-0

Appendix

Sequence Survey Questions

Leisure

Observing response:

You will now be asked to rank a list of activities in order of preference (1 = most preferred; 5 = least preferred). Before you do so, please complete the following statement:

"I understand I may not like any of these, but if I have to rank them, **of the activities listed**, the activity ranked #1 is my ...

- o MOST favorite (correct answer)
- o LEAST favorite

Q. Rank (drag and drop) the following activities in order of preference (1 = most preferred; 5 = least preferred).

_____ reading a magazine

_____ watching YouTube®

_____ spending time with your phone

_____ playing a paper game like Sudoku or crossword puzzles

_____ playing a computer game

Q. Imagine you get to spend the next couple of hours doing all those activities. In what order would you like to do the following activities?

Select the order in which you would like to experience each activity (first, second, third, fourth, last),

	Reading a Magazine	Watching YouTube®	Spending time with your phone	Playing a paper game like Sudoku or crossword puzzles	Playing a computer game
First	◯	◯	◯	◯	◯
Second	◯	◯	◯	◯	◯
Third	◯	◯	◯	◯	◯
Fourth	◯	◯	◯	◯	◯
Last	◯	◯	◯	◯	◯

School Work

Observing response:

You will now be asked to rank a list of activities in order of preference (1 = most preferred; 5 = least preferred). Before you do so, please complete the following statement:

"I understand I may not like any of these, but if I have to rank them, **<u>of the activities listed</u>**, the activity ranked #1 is my ...

- o MOST favorite (correct answer)
- o LEAST favorite

Q. Rank (drag and drop) the following activities in order of preference (1 = most preferred; 5 = least preferred).

_____ Completing worksheets with 3- and 4- digit multiplication problems, without a calculator.

_____ Completing worksheets with long division problems, without a calculator.

_____ Typing handwritten notes onto a computer

_____ Reading a text and answering reading comprehension questions (like those in the SAT)

_____ Work on building a model car with plastic rods and connectors

Q. Imagine you have to spend the next few hours working on school tasks. In what order would you prefer to work on these tasks?

Select the order in which you would like to work on each task (first, second, third, fourth, last),

	Completing worksheets with 3- and 4- digit multiplication problems, without a calculator.	Completing worksheets with long division problems, without a calculator.	Typing handwritten notes onto a computer	Reading a text and answering reading comprehensio n questions (like those in the SAT)	Work on building a model car with plastic rods and connectors
First	○	○	○	○	○
Second	○	○	○	○	○
Third	○	○	○	○	○
Fourth	○	○	○	○	○
Last	○	○	○	○	○

Noxious Stimuli

Observing response:

You will now be asked to rank a list of activities in order of preference (1 = most preferred; 5 = least preferred). Before you do so, please complete the following statement:

"I understand I may not like any of these, but if I have to rank them, **<u>of the activities listed</u>**, the activity ranked #1 is my ...

- o MOST favorite (correct answer)
- o LEAST favorite

Q. Rank (drag and drop) the following activities in order of preference (1 = most preferred; 5 = least preferred).

_____ Submerging your hand in very cold water (1^{o} C, or approximately 33^{o} F)

_____ Having the skin on your inner wrist pricked with a small plastic pin (feels like forcefully touching the sharp end of a pushpin).

_____ Experiencing the equivalent of 4 pounds of pressure in your inner wrist.

_____ Having the end of a thick and stiff nylon thread pressed repeatedly against the skin of your inner wrist 30 times in 30 seconds.

_____ Having a warm probe (50° C, or 122° F) touch the skin of your inner wrist.

Q. Imagine you go to a neurological appointment for testing. You have to experience each of these unpleasant sensations. In what order would you prefer to experience them?

Select the order in which you would like to experience each unpleasant stimulus (first, second, third, fourth, last).

	Submerging your hand in very cold water (1^{o} C, or ~ 33^{o} F).	Having the skin on your inner wrist pricked with a small plastic pin (feels like forcefully touching the sharp end of a pushpin).	Experiencing the equivalent of 4 pounds of pressure in your inner wrist.	Having the end of a thick and stiff nylon thread pressed repeatedly against the skin of your inner wrist 30 times in 30 seconds.	Having a warm probe (50° C, or 122° F) touch the skin of your inner wrist.
First	◯	◯	◯	◯	◯
Second	◯	◯	◯	◯	◯
Third	◯	◯	◯	◯	◯
Fourth	◯	◯	◯	◯	◯
Last	◯	◯	◯	◯	◯

Exercise

Observing response:

You will now be asked to rank a list of activities in order of preference (1 = most preferred; 5 = least preferred). Before you do so, please complete the following statement:

"I understand I may not like any of these, but if I have to rank them, **of the activities listed**, the activity ranked #1 is my ...

- o MOST favorite (correct answer)
- o LEAST favorite

Q. Rank (drag and drop) the following activities in order of preference (1 = most preferred; 5 = least preferred).

_____ Doing jumping jacks

_____ Doing sit-ups

_____ Doing push-ups

_____ Running in place

_____ Doing squats

Q. Imagine you have to do all these for 1 minute each, in an exercise session. In what order would you prefer to do these exercises?

Select the order in which you would like to do these exercises (first, second, third, fourth, last).

	Jumping jacks	Sit-ups	Push-ups	Running in place	Squats
First	◯	◯	◯	◯	◯
Second	◯	◯	◯	◯	◯
Third	◯	◯	◯	◯	◯
Fourth	◯	◯	◯	◯	◯
Last	◯	◯	◯	◯	◯

Food

Observing response:

You will now be asked to rank a list of foods in order of preference (1 = most preferred; 5 = least preferred). Before you do so, please complete the following statement:

"I understand I may not like any of these, but if I have to rank them, **of the foods listed**, the food ranked #1 is my ...

- o MOST favorite (correct answer)
- o LEAST favorite

Q. Rank (drag and drop) the following foods in order of preference (1 = most preferred; 5 = least preferred).

_____ Strawberries

_____ Grapes

_____ Cucumber

_____ Tomatoes

_____ Chips

Q. Imagine you have to eat all those foods. In what order would you like to eat them?

Select the order in which you would like to eat each food (first, second, third, fourth, last).

	Strawberries	Grapes	Cucumber	Tomatoes	Chips
First	◯	◯	◯	◯	◯
Second	◯	◯	◯	◯	◯
Third	◯	◯	◯	◯	◯
Fourth	◯	◯	◯	◯	◯
Last	◯	◯	◯	◯	◯

Mixed

Observing response:

You will now be asked to rank a list of activities in order of preference (1 = most preferred; 5 = least preferred). Before you do so, please complete the following statement:

"I understand I may not like any of these, but if I have to rank them, <u>**of the activities listed**</u>, the activity ranked #1 is my ...

- o MOST favorite (correct answer)
- o LEAST favorite

Q. Rank (drag and drop) the following activities in order of preference (1 = most preferred; 5 = least preferred).

_____ [Leisure #1]

_____ [School Work #1]

_____ [Noxious stimulus #1]

_____ [Food #1]

_____ [Exercise #1]

Q. Imagine you have to do or eat all the things listed above. In what order would you like to do or eat them?

Select the order in which you would like to do or eat each (first, second, third, fourth, last).

	[Leisure #1]	[School Work #1]	[Noxious stimulus #1]	[Food #1]	[Exercise # 1]
First	○	○	○	○	○
Second	○	○	○	○	○
Third	○	○	○	○	○
Fourth	○	○	○	○	○
Last	○	○	○	○	○

Endnotes

---

[i] A preference for improving sequences is apparent across categorically-different outcomes including money (e.g., Magen, Dweck, & Gross, 2008), painful stimulation (e.g., Ariely & Carmon, 2000), annoying sounds (e.g., Ariely & Zauberman, 2000), restaurant meals (e.g., Loewenstein & Prelec, 1993), health outcomes (e.g., Chapman, 2000), affective stimuli (e.g., Drolet, Lau-Gesk, & Scott, 2011), and news (e.g., Marshall & Kidd, 1981). Although the separate studies varied in multiple aspects including outcome category, sequence size, whether outcomes were hypothetical or real, and whether the sequence evaluations were conducted prospectively or retrospectively, some studies explicitly evaluating methodological variations (e.g., using allocation and pricing to determine preference, instead of order) have found that negative time preference in sequences may be related to the way in which the questions and sequences are framed (Frederick & Loewenstein, 2008). Therefore, variations in the way the task is framed may evoke different response tendencies.

Choice and decision-making researchers have formulated constructs rich in face validity to explain, or at least describe, why one may choose to forgo the immediate satisfaction of a highly valuable reinforcer. These include savoring and dread. Here, a formulation true to source material is provided and substantiated with apropos behavior analytic translations.

Loewenstein (1987) defined *savoring* as the process of deriving positive utility from the anticipation of desirable future outcomes and *dread* as the emotional impact of anticipating or contemplating future undesirable experiences. Alternatively, *savoring* may be conceptualized as engaging in overt or covert behavior associated with conditioned positive reinforcement related to an upcoming outcome. Similarly, *dread* can be conceptualized as overt or covert behavior associated with conditioned aversive stimuli related to the upcoming outcome. Taken together, the construct of anticipation, and more specifically, savoring and dread, suggest that simply anticipating positive experiences is pleasurable, and anticipating negative experiences is inherently aversive. For example, one might derive pleasure and excitement in contemplating an upcoming vacation, or fear and dread when thinking of an upcoming dissertation defense.

Economists and social scientists have recognized the importance of anticipation as a source of positive or negative emotions for some time. In a study of health records of factory employees in the 2 years preceding a factory closure, the greatest number of illnesses were recorded in the period when unemployment was anticipated rather than during actual unemployment (Kasl, Gore, & Cobb, 1975). In another study, when asked to rank days of the week based on preference, college students who did not have classes on Saturday ranked Sunday lower than Friday. This was despite classes being held on Friday, making Fridays objectively part of the school-week and Sundays part of the weekend (Farber, 1953). Based on participants' comments, Farber interpreted the higher ranking of Friday over Sunday to mean that the anticipation of the weekend made Fridays more enjoyable, whereas the anticipation of the work- or school-week made Sundays more dreadful.

Psychologists have used the term *savoring* to describe the self-regulation of positive feelings, sensations, perceptions, thoughts, and behaviors one engages in when attending to and appreciating a positive stimulus (Bryant, Chadwick, & Kluwe, 2011).

Bryant et al. (2011) suggested several savoring processes underly positive experiences. For example, in addition to savoring a present experience, one may also derive positive feelings and sensations from savoring a past (i.e., reminiscence) or a future (i.e., anticipation) experience, or one may even enhance the quality of a present positive experience by remembering looking forward to the experience (i.e., recalled anticipation) or by looking forward to reminiscing about the experience at a later time (i.e., anticipated recall). Research on episodic memory (Tulving, 1984), episodic future thinking (Atance & O'Neill, 2001), and "mental time travel" (Quoidbach, Wood, & Hansenne, 2009; Suddendorf & Corballis, 2007) provides support for the idea that humans have the capacity to re-experience a past event or pre-experience a future one. In other words, by imagining future or past positive events one can experience positive feelings in the present and one may enhance the emotional perception of a current event by reflecting on it from a past or future-focused perspective.

Thus, *savoring* and *dread* may promote preference for improvement because for preferred outcomes, improving sequences allow the individual to produce conditioned positive reinforcers, or savor the best outcome until the end (Loewenstein & Prelec, 1993). Conversely, for aversive experiences, improving sequences allow the individual to experience the most aversive outcome first, in a way "getting it over with," therefore reducing the amount of dread to be experienced.

[ii] In addition to framing and outcome category, there are many individual variables that might influence one's time preference and decision making. For example, the behavioral inhibition and behavioral approach systems are thought to influence sensitivity to

reinforcement and punishment and individual differences in sensitivity to these physiological self-regulatory systems can predict level of reported nervousness in response to an impending noxious experience and happiness in response to an impending rewarding experience (Carver & White, 1994). As such, individual differences in behavioral inhibition and behavioral approach systems sensitivities may also predict individual differences in the types of sequences participants generate. Similarly, anhedonia, a decreased reactivity to pleasurable stimuli that is a characteristic feature of depression (Admon & Pizzagalli, 2014), might influence the type of sequence someone might generate. Likewise, mindfulness, or specifically present-focus, may also influence the type of sequence one generates. A person who is focused in the moment, may be less likely to engage in private events that are associated with conditioned positive reinforcement or conditioned aversive stimuli related to upcoming outcomes.

[iii] For example, if a participant ranked the following foods from most to least preferred as follows: strawberries (1^{st}), chips (2^{nd}), crackers (3^{rd}), tomatoes (4^{th}), and cucumbers (5^{th}), and then proceeded to generate a sequence in the following order: cucumbers, tomatoes, crackers, chips, and strawberries, then the Spearman rank correlation for this participant would be -1.

[iv] The overall BIS-11 scale and the attentional impulsiveness subscale showed good reliability (i.e., $\alpha \geq .7$; Nunnally & Bernstein, 1994), indicating high correlation among the different items. The reliability estimate dropped slightly for the non-planning and motor impulsiveness subscales.

---

[v] Materials for the work category included writing utensils such as pens or pencils, worksheets with three- and four-digit multiplication problems, worksheets with long-division problems, a computer with access to Microsoft Word, a copy of a textbook chapter, reading comprehension worksheets, and a K'Nex® model building set. Materials for the food category included disposable plates, napkins, and disposable gloves for handling food. Additionally, five foods that were pre-approved by the participants at the time of enrollment: apples, grapes, carrots, celery and hummus, and popcorn. Materials in the leisure category included magazines, a tablet with access to YouTube®, the participant's cell phone, paper copies of Sudoku and crossword puzzles with writing utensils such as pencils and pens, and a computer with access to online games. Materials for the noxious category included a stainless-steel bucket with iced-water, a warm thermal probe, an algometer, a Von Frey monofilament (60 g), and a single-use neurological pinprick sensory testing tool. Materials for the exercise category included a floor mat and picture cards depicting the different exercises. Materials for the mixed category consisted of select materials from each of the other categories.

[vi] **Precautions and participant protection.**

Consent. During the consent meeting, participants were informed of health-related matters that could place the individual in harm in a cold-pressor task, including pulmonary hypertension, history of high blood pressure, diabetes, stroke, fainting, and Raynaud's disease. Before participants signed the consent form, they were asked to complete the Evaluation to Sign Consent, which consisted of several questions to ensure participants understood the consent form and research protocols. Upon correct completion of this evaluation, the consent forms were signed. Any participants who reported having any of the aforementioned health conditions would have been excluded from the noxious and mixed categories but could continue to participate in the remainder of the study, if they choose to do so. This was never the case, however, as none of the participants reported having any of those health conditions. Following consent, participants were asked to select five foods from a list and note any potential allergies or dietary restrictions. They were also asked to bring their cell phone to session and to dress comfortably enough to engage in light physical exercise.

Experimenter training. Experimenters were trained, prior to conducting sessions, on how to safely implement the procedures involved in the noxious category as is conducted in research involving Modified Quantitative Sensory Testing. Training involved discussion and role play as participant and experimenter.

Modifications from acceptable practices. Exposure to the noxious stimuli in this study was never more than what is accepted for use in research settings. For example, the cold-pressor task in this study was limited to 5 s, even though trials of as much as 3 to 5 min are considered acceptable with children and adults (Mitchell, MacDonald, & Brodie, 2004; von Baeyer, Piira, Chambers, Trapanotto, & Zeltzer, 2005).

[vii] If a participant were to refuse to experience a stimulus, data collectors would note the refusal, and the experimenter would proceed to the next trial. If a participant were to refuse to experience all stimuli, this participant's data would not be included in the analysis for the noxious category, and the participant would not be able to experience the mixed category because it also involved one of the noxious stimuli. To be included in the

analysis, participants had to experience at least 80% or 4/5 of the outcomes in the array. None of the participants refused to experience any stimulus, so all experienced 100% of the outcomes in the array.

[viii] If the participant elected to not consume the item, the food was placed on a "Discard" plate to the side. The experimenter would then present the next item. These steps continued until the participant had sampled all foods in the selected order. If a participant refused to experience all stimuli, this participant's data would not be included in the analysis for the food category, and the participant would not be able to experience the mixed category because it also involved one of the foods. To be included in the analysis, participants had to experience at least 80% or 4/5 of the outcomes in the array. None of the participants refused to consume any of the foods, so each experienced 100% of the outcomes in the array.

[ix] The items in the work category included activities that college students should be able to do (e.g., transcribing a textbook chapter into a Word document, reading comprehension work-sheets, long-division and multiplication, and building a K'Nex model that is appropriate for elementary-school children), therefore participants were not given assistance throughout the task. Nevertheless, the K'Nex model building task had an instruction/model sheet, and written instructions on how to do long multiplication and division were made available to all participants along with the task materials.

[x] As the last part of the study, participants were asked to complete an exit survey with three questions. In the first open-ended question, participants were asked why or how they chose the order in which they wanted to experience each series of items and whether

anything specifically influenced their choice. In the second question participants were asked to rate each item they experienced using a Likert-type scale with options: love, like, neutral, dislike, and hate. In the third question, they were asked to choose the

description that best described their perception about their general mindset, among three options: (a) generally focused on the present, (b) equally focused on the present and the future, and (c) generally preoccupied by the future.

[xi] Cohen's kappa (McHugh, 2012) was also calculated as an additional estimate of IOA for activity engagement. There was fair agreement between the two observers for Cole's engagement, $\kappa = .38$ ($p < .001$, 95% CI [.18, .59]). There was substantial agreement for Andrea $\kappa = .66$ ($p < .001$, 95% CI [.23, 1.10]) and David $\kappa = .69$ ($p < .001$, 95% CI [.57, .81]), and perfect agreement for Carla $\kappa = 1.00$ ($p < .001$).

[xii] Kahneman et al. (1993) exposed participants to two aversive experiences: in a short trial, participants submerged one hand in cold water (14°C) for 60 s; in the long trial, participants submerged the other hand in cold water (14°C) for 60 s and then kept the hand submerged for an additional 30 s as the water temperature gradually increased by 1°C (i.e., an experience that became slightly less aversive toward the end). When participants were asked next to choose which trial they preferred to repeat, 68% of participants chose to repeat the long trial. Kahneman et al. concluded that participants' preference for more pain over less was due to retrospective evaluations of aversive experiences being influenced by the peak and end moments of the experience.

Save the Best for Last III: Some Children Save the Best for Last in the MSWO

Mariana I. Castillo[1], Michelle A. Frank-Crawford[1&2], Jody E. Liesfeld[3], Trang M. Doan[3], Eli T. Newcomb[3], Griffin W. Rooker[2&4], and John C. Borrero[1]

UMBC[1], Kennedy Krieger Institute[2], The Faison Center[3], and Johns Hopkins University School of Medicine[4]

Completion of this study and preparation of this manuscript were supported in part by a grant from the UMBC Graduate Student Association.

Mariana Castillo ORCID 0000-0002-7433-4324

Michelle Frank-Crawford ORCID 0000-0003-4319-6317

Griffin Rooker ORCID 0003-0246-0775

John Borrero ORCID 0000-0003-0413-0892

Direct correspondence to J. C. Borrero, Department of Psychology, UMBC, 1000 Hilltop Circle, Baltimore, Maryland 21250 (e-mail: jborrero@umbc.edu).

Abstract

The Multiple Stimulus Without Replacement (MSWO; DeLeon & Iwata, 1996) is commonly used to identify potential reinforcers. In this arrangement, all items are presented in an array, and the individual is prompted to "pick one." After consuming or engaging with the item for some time the remaining items are presented in the next trial. Presumably, the item selected first is the highest preferred, and the item selected last is the least preferred. Scattered reports suggest that some individuals may save the best for last (STBFL) in an MSWO. No explicit studies of the phenomenon have been conducted so far. We aimed to determine whether individuals with intellectual and developmental disabilities STBFL in an MSWO. The stimuli ranked first and last in the MSWO were evaluated as reinforcers under progressive ratio schedules. Based on the mean break point, one of four participants STBFL in the MSWO. Results were consistent for edibles and toys.

Keywords: MSWO, save the best for last, preference for sequences, negative time preference, individuals with intellectual and developmental disabilities.

Save the Best for Last III: Some Children Save the Best for Last in the MSWO

Stimulus preference assessments are often conducted to identify preferred stimuli to use as reinforcers in behavioral interventions for skill acquisition or the reduction of challenging behavior. One method for identifying a preference hierarchy is the multiple-stimulus without replacement (MSWO) preference assessment (DeLeon & Iwata, 1996). The MSWO is considered one of the most time-efficient preference assessment procedures, which allows practitioners to identify a hierarchy and multiple potential reinforcers in minimal time (Karsten, Carr, & Lepper, 2011). Unlike some other stimulus preference assessments, the MSWO involves presenting an array of several items to the individual at once; therefore, this procedure is recommended for individuals who can scan an entire array and select a preferred item without exhibiting a positional bias, such as selecting the item that is on the right every time (Chazin & Ledford, 2016; Karsten et al., 2011). At the conclusion of the assessment, the item selected first has the highest selection percentage and is therefore the highest ranked item, and presumably the most preferred.

Selection of items from most-to-least preferred can be explained by a general preference for immediacy. When given a choice, humans generally prefer tangible or monetary rewards sooner rather than later (Loewenstein, 1987; Odum, 2011). This preference for immediacy, or conversely, aversion to delays, is supported by the extensive research on temporal discounting, which has shown that in choice paradigms where only one outcome is selected from the array, delayed outcomes are generally valued less than those without a delay (Ainslie, 1975). In economics, this is referred to as *positive time preference* (Olson & Bailey, 1981). If positive time preference is exhibited when selecting items from an array, then the highest preferred item should in fact be the one selected first during an MSWO.

Nevertheless, there are reports that suggest that at times, individuals selecting items in an MSWO may be saving the best for last (Soldberg, Hanley, Layer, & Ingvarsson, 2007). In a study of the correspondence between mean preference rank and performance during a reinforcer assessment, Call, Trosclair-Lasserre, Findley, Reavis, and Shillingsburg (2012) found that, for one of the seven participants (i.e., Cameron), the item ranked highest in the paired-stimulus preference assessment had the lowest rank in an MSWO, and was also associated with the highest mean break point during the reinforcer assessment. Although not discussed in these terms, these data suggest that it is possible that Cameron was saving the best for last in the MSWO.

Saving the best for last is an example of *negative time preference*, or a general preference for improvement (Loewenstein & Prelec, 1991). If individuals are exhibiting negative time preference and choosing to save the best for last in an MSWO, then practitioners may not identify the highest preferred item. Not identifying the most preferred item may not always be a problem, given that some studies have found that lower preferred items may also function as reinforcers, particularly when tested independently (e.g., Francisco, Borrero, & Sy, 2008). Nevertheless, preference assessment outcomes often correspond with the outcomes of reinforcer assessments, such that higher preferred items support more responding (DeLeon, Frank, Gregory, & Allman, 2009; Glover, Roane, Kadey, & Grow, 2008). In cases where preference is correlated with reinforcer efficacy, misidentifying the highest preferred item may result in the selection of a less efficacious reinforcer. In some cases, practitioners may use more than one reinforcer. Using a variety of reinforcers has been a solution to minimize the detrimental effects of satiation on subsequent reinforcer efficacy (Keyl-Austin, Samaha, Bloom, & Boyle, 2012). If reinforcer variation is to be used, items identified as *high*- and *moderately*-preferred are likely to be included (Keyl-Austin et al., 2012) and not necessarily those identified as *low* preferred. If an individual saves the best for last in an MSWO, however, the item selected last (presumably the least preferred) is in fact the highest preferred, in which case practitioners may be missing out on the most effective reinforcers.

The prevalence of saving the best for last in MSWOs is unclear, however, and may be masked by publication bias (Tincani & Travers, 2019). It is possible that researchers may switch to a different assessment method (e.g., paired-stimulus preference assessment) if participant responding does not conform to what is expected in an MSWO. Then, when published, the final manuscript may only reflect the final preference assessment conducted. As mentioned previously, there is one known publication in which a participant was explicitly noted to be saving the best for last (Soldberg et al., 2007), all other reports come from conference proceedings (Becerra & Fahmie, 2014; Litchmore, Ivy, & Weaver, 2014; Ngur, Dillon, & Bowman, 2018; Pendharkar, Bourret, Nuzzolilli, & Upshaw, 2017; Roath & Fritz, 2015), anecdotal reports, or our own clinical practice. Given the limited published evidence of "saving the best for last" in an MSWO, it is not surprising that there are currently no published explicit studies of the phenomenon. When saving the best for last does occur, verbal behavior may be implicated.

Human choice has long been related to language[xiii], or verbal behavior[xiv]. Studies involving concurrent choice arrangements have suggested that there are complex interactions between experimental contingencies, the participants' verbal behavior, and participants' performance (Harzem, Lowe, & Bagshaw, 1978; Horne & Lowe, 1993). Furthermore, research on time preference has suggested that when the choice is *framed* as part of a sequence, people tend to prefer saving the best for last (Castillo, Sun, Frank-Crawford, & Borrero, 2020; Loewenstein & Prelec, 1991, 1993). In a brief survey, Loewenstein and Prelec (1991) asked college students to choose between dinner at their highest preferred restaurant in one month or in two months. The majority (80%) selected the sooner option, exhibiting positive time preference. When the participants were later given the choice of scheduling two dinners: one at a moderately preferred restaurant and one at a more preferred restaurant, 57% of participants elected to experience the moderately preferred meal in one month and the more preferred meal in two months (i.e., they exhibited negative time preference; Loewenstein & Prelec, 1991). In other words, participants chose to delay the highest preferred outcome when the choice was framed as a sequence. One possibility is that the MSWO arrangement might be perceived as an opportunity to select the order in which a sequence of outcomes is experienced. If the framing of the choice (e.g., selection of a single-outcome choice versus scheduling a sequence of unavoidable outcomes) does in fact influence how people respond, then perhaps saving the best for last is related to language ability.

In other research, language ability has been positively associated with delay of gratification among impulsive children (Rodriguez, Mischel, & Shoda, 1989). Longitudinal studies have suggested that receptive language and vocabulary predict later behavior problems such as inattention-hyperactive and externalizing problems (Peterson et al., 2013). Receptive language has been positively associated with delay of gratification by children with Down syndrome (Cuskelly, Gilmore, Glenn, & Jobling, 2016). Psychologists have theorized that language skills may enable executive control and meta-cognitive processing by facilitating self-reflection, response inhibition, and behavioral direction (Gallagher, 1999). Although self-control as measured in the cited research is not equivalent to preference for improving sequences, it is still possible that language ability is also related to an individual’s choice to save the best for last in an MSWO preference assessment.

As noted previously, although there are some anecdotal reports of individuals saving the best for last in MSWO arrangements, no explicit studies of the phenomenon have been conducted. The purpose of this study was to determine whether individuals with intellectual and developmental disabilities (IDD) who have moderate or high verbal ability would save their most preferred item for last in an MSWO preference assessment. Following the MSWO, items selected first and last were then assessed under progressive ratio schedules to determine reinforcer efficacy.

Method

Participants and Setting

A total of five children with IDD were recruited from two different schools for children with autism spectrum disorder (ASD) and other developmental disabilities and were identified by teachers or other school personnel as having moderate or high verbal ability. Jackson was 9 years and 10 months old at the start of the study and had a diagnosis of ASD. David was 15 years and 3 months and was diagnosed with ASD and epilepsy. Connor was 9 years and 7 months old and was diagnosed with Smith–Magenis Syndrome. Laura was 9 years and 11 months old at the start of the study and had a diagnosis of ASD. Carlos was 12 years and 6 months and diagnosed with ASD and attention deficit hyperactive disorder. During baseline sessions, Carlos continuously worked at high rates, and even requested to continue working following the end of session. Even after increasing task difficulty, Carlos continued working uninterruptedly during baseline, therefore his participation was discontinued as we did not believe that we could demonstrate a reinforcement effect. Only the four participants who completed the entire study will be discussed hereafter. Sessions were conducted in a quiet room in the school, such as an office or staff lounge, furnished with two chairs and a table.

Language Ability Measure

To get an objective measure of language ability, beyond teacher report, all participants completed the Peabody Picture Vocabulary Test, fourth edition (PPVT-4; Dunn & Dunn, 2007). The PPVT-4 is a test of receptive vocabulary that is individually administered and provides an estimate of verbal ability. The test is given vocally, it does not require spoken language from the test taker and takes 10–15 min to administer. For its administration, the examiner presents a series of pictures (four images to a page) and asks the participant to point to the picture that corresponds to a specific word. The total score can be converted to an age-normed standard score with a mean of 100 and standard deviation of 15, a percentile rank, or an age-equivalent. Valuation studies (Dunn & Dunn, 2007) indicated strong psychometric properties, with split-half reliability and Cronbach's alpha of 0.94 and test-retest reliability of r = 0.93. The PPVT-4 is considered a valid measure of receptive language ability. It was used as a measure of convergent validity for the Toolbox Picture Vocabulary Test developed as part of the NIH Toolbox Cognition

Battery (Gershon et al., 2013). It was also recently identified as a valid proxy for verbal IQ in large-scale studies of ASD (Krasileva, Sanders, & Hus Bal, 2017).

Procedures

All participants completed a delay sensitivity assessment first, followed by preference and reinforcer assessments with edibles and leisure items separately. The order was randomized across participants such that some participants completed the assessments with edibles first (i.e., Jackson and Connor), and others completed them with leisure items first (i.e., David and Laura).

Delay sensitivity assessment. Participants completed a brief delay sensitivity assessment. Specifically using an item identified by the caregivers as highly preferred, the experimenter told the participant: "We have a [preferred item] for you. Do you want this [preferred item] now or later, in about 3 minutes?" If the participant selected "now," the item was delivered immediately. If the participant selected "later" the experimenter started a 3-min timer and said "OK, you chose in 3 min. You can have your [preferred item] when the timer beeps." When the 3-min timer elapsed, the experimenter delivered the item to the participant. The procedure was repeated two more times, for a total of three trials.

MSWO preference assessments. Five items that could be presented on a table top and were previously identified by the participants' caregivers as highly preferred were included. The MSWO preference assessment was completed as described by DeLeon and Iwata (1996), with the exception that as few as three sessions, and up to five sessions, were conducted. A recent study found that for MSWOs of edible or leisure items, three sessions were significantly and positively correlated with the outcomes from the 5-session MSWOs for all participants (Richman, Barnard-Brak, Abby, & Grubb, 2016). All participants completed the MSWO a minimum of three times. If the items ranked first and last according to the mean of all sessions, matched the items ranked first and last in the last session, no more sessions were conducted. If the items differed, then an additional session was conducted, up to a maximum of five sessions. In the MSWO all stimuli were evenly spaced on the table in a horizontal array. Prior to each trial, the experimenter established eye contact with the participant and prompted the participant to orient to all stimuli ("[name], look") by making a sweeping gesture from one end of the array to the other. The experimenter then prompted the participant to pick one. If the participant did not select an item after 10 s, the prompt to pick one was repeated. The selected item was delivered immediately following selection to engage with for 30 s (leisure items) or to be consumed (edibles). After the access period elapsed, or the participant had finished consuming the edible, the selected item was either removed from the immediate area (leisure item) or not replaced (edible). Attempts to select more than one stimulus per trial were blocked. Between trials, the sequencing of the remaining items was rotated by taking the item from the left side of the array and moving it to the right end, then shifting the other items so that they were again equally spaced on the table. These steps were repeated until all stimuli were selected or no stimulus was selected within 30 s from the beginning of a trial.

Overall ranks were determined by adding the number of times each item was available and assigning the highest rank (1) to the item with the lowest sum (see also Karsten, Carr, & Lepper, 2011). Although percentage of selections is commonly used to determine ranks in the MSWO, using sum of ranks to determine overall ranks seemed more appropriate given the experimenters' interest in ordinal selection. The items with the highest and lowest overall rank were selected for the reinforcer assessment.

Reinforcer assessments. Following the teacher's recommendation, the experimenters selected a task that the participants could complete at their own rate. Jackson and Connor traced letters, and David did single-digit addition and subtraction. Laura did simple addition in the first assessment (leisure items), but given the low levels of responding, switched to tracing letters for the second assessment (edibles). To ensure that engaging in the task was not inherently reinforcing, a brief no-reinforcement baseline was conducted first. The stimuli ranked highest and lowest in the MSWO were evaluated as reinforcers under progressive ratio (PR) schedules presented in single-operant arrangements.

Baseline. A minimum of three baseline sessions were conducted. Session duration varied based on participants' performance. Prior to initiating a session, participants were prompted to complete one task. No consequences were provided following completion. The following instructions were provided: "When you [engage in the target response], you will not get anything. You can [emit the target response] if you want to, but you don't have to. If you ever want to stop the session you can tell me or hand me this stop sign." If the participant engaged in problem behavior or emotional responses (e.g., crying or screaming) at any point during the session, the experimenter repeated the session instructions. Sessions were terminated immediately following the first instance of any of these criteria: (a) a 1-min period without a target response, (b) a withdrawal of assent from the participant (i.e., a request to stop that could be either vocal, or pointing to a picture of a stop sign that was placed at the side of the table), or (c) 30 min of session time.

Reinforcement. Three PR session, at minimum, were conducted with each stimulus (items with highest and lowest rank in the MSWO). Reinforcement sessions were conducted in a multielement design, alternating stimuli across sessions in a randomized order. Session duration varied based on participants' performance. Only one PR session was conducted per meeting. Meetings were conducted once per day or twice per day, with at least a 1-hr break in between meetings. The reinforcement sessions with a given stimulus were terminated when responding met stability with that item, defined as three consecutive sessions in which the number of schedules completed did not differ by more than three and there was no observable trend in the data, or when a maximum of six reinforcement sessions were conducted per stimulus.

Prior to initiating session, participants were prompted to complete the task (guided exposure, according to the smallest schedule requirement). The session-specific stimulus was provided following completion of the schedule requirement. Following guided exposure, the session instructions were provided: "When you [engage in the target response], you will get [specific stimulus]. You can [emit the target response] if you want to, but you don't have to. If you ever want to stop the session you can tell me or hand me this stop sign." The session-specific stimulus was delivered following completion of the pre-specified schedule requirement. The number of responses required to produce a reinforcer increased throughout each session according to an arithmetic progression with an addition of a fixed number of responses. That is, following each reinforcer delivery, the response requirement to produce the next reinforcer was increased by a constant number until one of the stop criteria was reached. For Jackson, the staring response requirement (SRR) was 1, with a PR step size of 2 (i.e., PR 2). David's response requirement was set to SRR 2, PR 2. Connor's response requirement was set to SRR 5, PR 5 for both assessments. Laura's response requirement was set to SRR 1, PR1 during the assessment of leisure items, and SRR 5, PR 5 during the assessment of edibles. If at any time during the session the participant engaged in problem behavior or emotional responses (e.g., crying or screaming), the experimenter repeated the session instructions. Sessions were terminated immediately following the first instance of any of these criteria: (a) a 1-min period without a target response, (b) a withdrawal of assent from the participant (i.e., a request to stop that could be either vocal, or pointing to a picture of a stop sign that was placed at the side of the table), or (c) 30 min of session time.

Data Collection and Interobserver Agreement

Observers used paper data sheets to record the order in which each item was selected during the MSWO. Interobserver agreement (IOA) was calculated for each trial, and total agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements across all trials. With the exception of David's and Connor's MSWO of edibles items, for which only 50% and 60% of trials had IOA, all other MSWOs had IOA calculated for all trials. There was 100% agreement on the order each item was selected in all trials scored by two observers.

During the reinforcer assessments, the observers collected data on task completion, whether the participant completed the schedule requirement, reinforcer deliveries, and which session termination criterion applied. A second observer independently collected data during 69% (Jackson), 51% (David), 71% (Connor), and 74% (Laura) of sessions, with a minimum of 33% (Jackson), 25% (David), 25%

(Connor), and 50% (Laura) in each condition and phase. For baseline task completion, schedule requirement and reinforcer deliveries, total count IOA was calculated by dividing the smaller number of responses by the larger number of responses for each session (Reed & Azulay, 2011). These fractions were averaged across all sessions with two observers to obtain the percentage of agreement. For task completion during reinforcement, trial-by-trial IOA was calculated by counting the number of schedule values for which there was agreement in the number of responses completed and dividing the number of agreements by the number of agreements plus disagreements (Reed & Azulay, 2011). For session termination criteria, exact agreement was used to determine IOA. Mean total count IOA for task completion in baseline was 98.76% (range, 92.59 – 100) for Connor, and 100% for Jackson, David, and Laura. Mean trial-by-trial IOA for task completion in reinforcement was 97.08% (range, 86.67 – 100) for Jackson, 97.17% (range, 85.71 – 100) for Connor, and 100% for David and Laura. Perfect agreement (100% IOA) was obtained for all participants in all other measures.

Results

Results of the receptive language measure, the PPVT-4, are depicted in Table 1. All participants spoke in full sentences. During the delayed sensitivity assessment, all participants chose to receive the item immediately, rather than after 3 min. In other words, when given the choice between receiving a reward immediately, or the same reward following a delay, participants preferred the immediate outcome, exhibiting *positive time preference* (i.e., preference for present over future consumption; Olson & Bailey, 1981).

Table 1

Results of the Peabody Picture Vocabulary Test

		PPVT Score	
Participant	Age at test date	Standard Score (CI)	Age Equivalent
Jackson	9:10	74 (67-83)	6:7
David	15:6	43 (38-51)	5:11
Connor	9:7	61 (55-70)	5:2
Laura	9:11	70 (64-79)	6:1

Note. Ages are listed as years and months (y:m). PPVT = the Peabody Picture Vocabulary Test, fourth edition. A standard score indicates the distance of the participant's raw score from the mean for people of the same age. A standard score of 100 is the average score for the person's age. The standard deviation for the PPVT-4 standard scores is 15. CI = 95% confidence interval. An age equivalent represents the age at which a participant's raw score is the mean score in a growth curve across age.

Results of the MSWO preference assessments are depicted in Table 2. The items ranked first and last were selected for the reinforcer assessments. David never selected Twix and Laura never selected Graham Bunnies in the MSWO; therefore, the items they *selected* last, ranked #4 (i.e., Goldfish and PopChips, respectively), were included in their reinforcer assessments.

Table 2

Overall Rank of Each Stimulus in the MSWO Preference Assessments for Each Participant

			Overall Rank Across all Sessions				
Participant	Stimuli	Number of Sessions	1	2	3	4	5
Jackson	Edibles	3	Veggie Straws	Chips	M&M's	Chocolate-covered Pretzels	Fruit Snacks
	Leisure	3	Hula Hoops	Slinkies	Bubbles	Play Doh	Slime
David	Leisure	3	Slime	Spinning Top	Squeeze Ball	Slinky	Rubix Cube
	Edibles	4	Chip & French Onion Dip	Popcorn	Chip & Ranch Dip	Goldfish	Twix Bar (NS)
Connor	Edibles	5	Chips Ahoy	Oreos	Fruit Snacks	Veggie Straws	PopChips
	Leisure	3	Slime	Slinkies	Spikey Ball	Bubbles	Play Doh
Laura	Leisure	3	Tablet	Shimmer & Shine Toy	Coloring Book	Water Hoop Game	Jewelry
	Edibles	5	Fruit Snacks	Utz Chips	Ginger Snaps	PopChips	Graham Bunnies (NS)

Note. NS = Never Selected.

Results of the reinforcer assessments are discussed next and are presented in separate figures for each participant. Within each figure, the assessments are presented in the order they were conducted. That is, the assessment presented in the top panel of the figure was conducted first, and the assessment in the bottom panel was conducted last. The left panels depict the total responses completed in each session of the reinforcer assessments. The right panels depict the break point analyses. Given the established stability criteria (i.e., three consecutive sessions in which the number of schedules completed for a given stimulus did not differ by more than three and there was no observable trend in the data), the experimenters used the data for the last three sessions to calculate the mean break point for each stimulus.

Results of Jackson's reinforcer assessments are depicted in Figure 1. He completed the assessment with edibles first (top panels), and then with leisure items (bottom panels). In the first assessment (top left panel), he completed very few responses in baseline ($M = 1.66$). The number of responses completed per session increased when the stimulus ranked last was used as a reinforcer ($M = 449.75$), and initially increased, but then dropped, when the stimulus ranked first was used as a reinforcer ($M = 90.33$). The break point analysis (top right panel) indicates that the stimulus Jackson selected last in the MSWO had a higher break point ($M = 45$), than the stimulus he selected first ($M = 4.33$), which suggests that Jackson was saving the best for last in the MSWO of edibles. Similar response patterns were observed in the assessment of leisure items (bottom panels). He completed very few responses in baseline ($M = 14.66$). The number of responses completed per session increased when the stimulus ranked last was used as a reinforcer ($M = 245$).

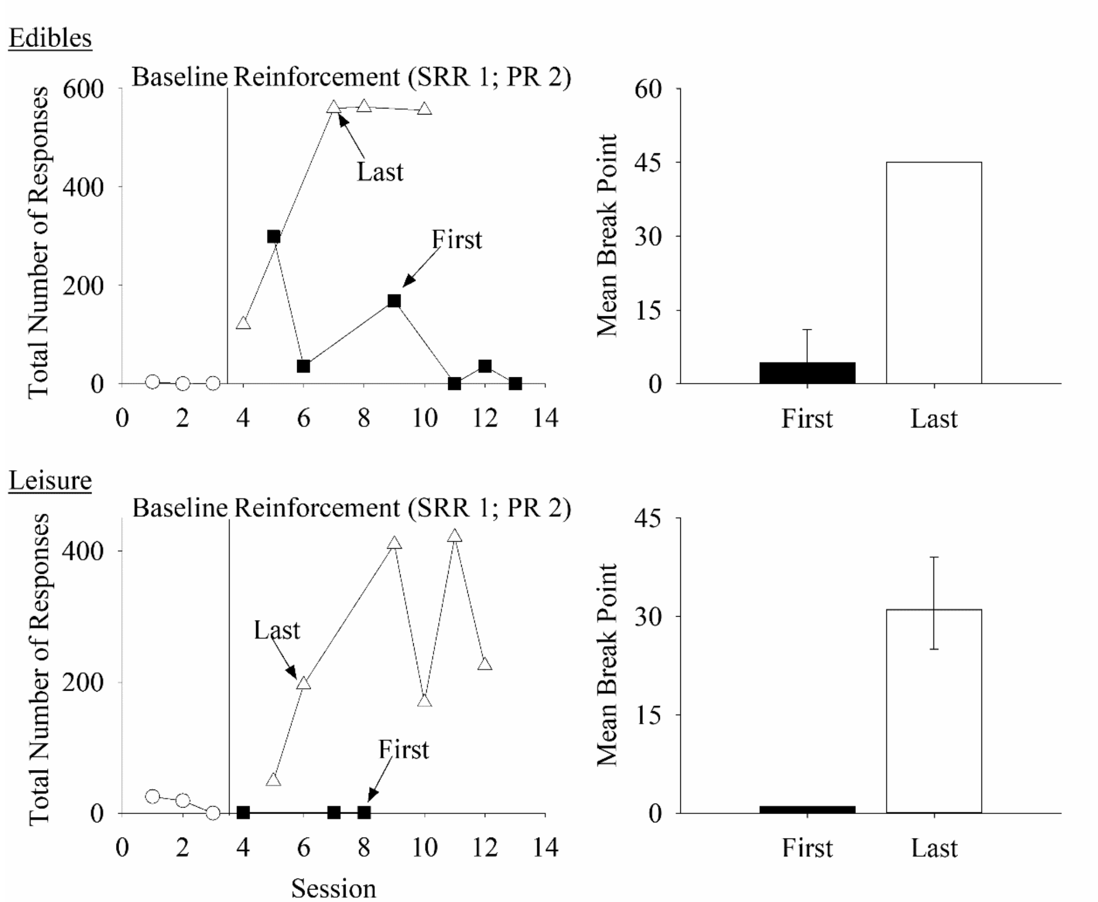

Figure 1. Results of Jackson's reinforcer assessments. Top panels depict the assessment of edibles and bottom panels depict the assessment of leisure items. In the left panels, each data point depicts the total number of correct responses in a specific session. The open circles are baseline sessions. The closed squares depict reinforcement sessions with the stimulus ranked first in the MSWO. The open triangles depict reinforcement sessions with the stimulus ranked last in the MSWO. The right panel depicts the break point analyses. The black bar depicts the mean break point (last schedule value completed) across the last three sessions for the stimulus ranked first in the MSWO. The white bar depicts the mean break point across the last three sessions for the stimulus ranked last in the MSWO. The error bars depict the range of break points for each stimulus. SRR = Starting response requirement. PR = progressive ratio step size.

Jackson completed very few responses when the stimulus ranked first was used as a reinforcer (M = 1), suggesting it did not function as a reinforcer. The break point analysis (bottom right panel) indicates that the stimulus Jackson selected last in the MSWO had a higher break point (*M* = 31), than the stimulus he selected first (*M* = 1), which suggests that Jackson was also saving the best for last in the MSWO of leisure items.

Results of David's reinforcer assessment are depicted in Figure 2. He completed the assessment with leisure items first (top panels), and then with edibles (bottom panels). In the first assessment (top left panel), task completion was initially variable in baseline, but ended in a decreasing trend (*M* = 11). Responding increased during reinforcement, with a slightly higher number of responses completed per session, on average, when the stimulus ranked first was used as a reinforcer (*M* = 48.83), than when the stimulus ranked last was used as a reinforcer (*M* = 46.40). Looking at the last three sessions with each stimulus, however, responding was higher and more stable with the stimulus ranked last, than it was with the stimulus ranked first. The break point analysis (right panel) includes data from the last three sessions with each stimulus. Results indicate that the stimulus David selected last in the MSWO had a slightly higher break point (*M* = 14.66), than the stimulus he selected first (*M* = 10.66), which suggest that David may have saved the best for last in the MSWO of leisure items. Nevertheless, this conclusion is made with caution, given the variability in David's responding during the assessment and the degree of overlap between both data paths. Furthermore, David's reinforcer assessment took a substantially greater amount of time to complete than did the same assessments for other participants (see Table 3). Considering the number of days required to complete David's reinforcer assessment, an alternative explanation for the difference in mean break points is that David's preference shifted throughout the course of the assessment.

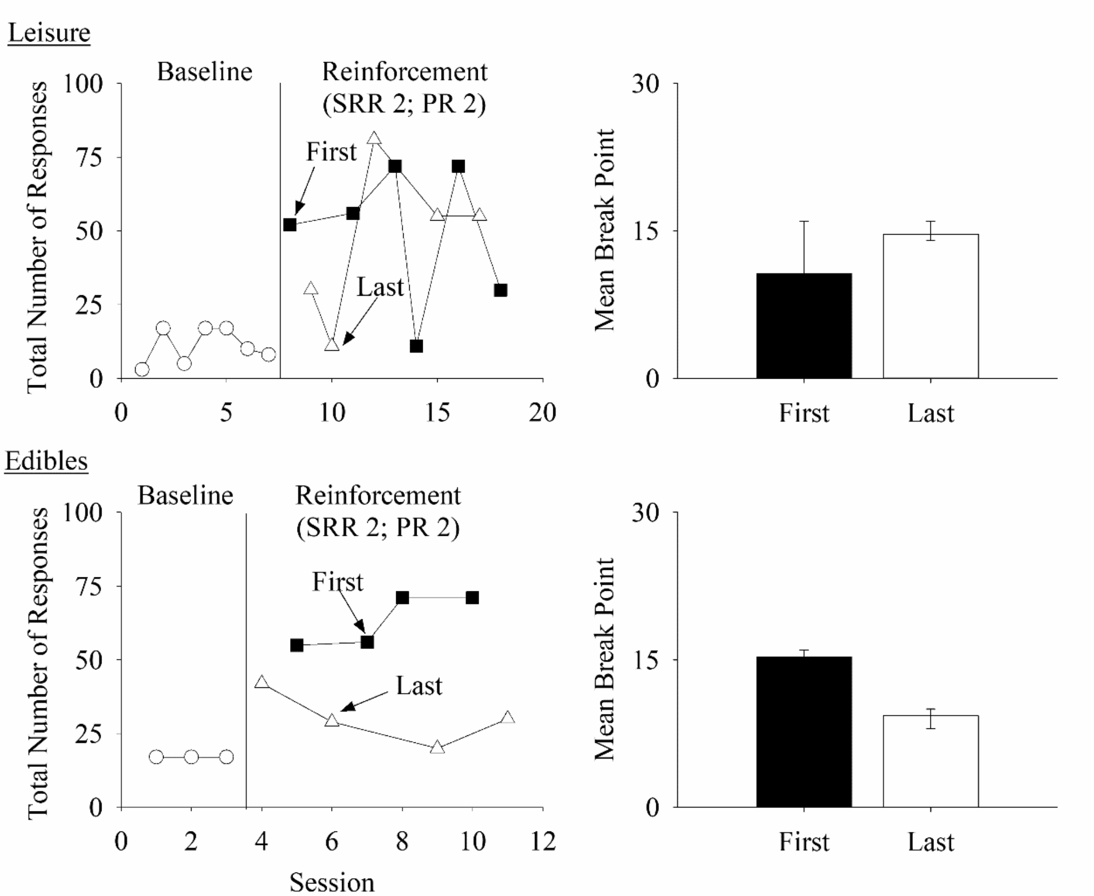

Figure 2. Results of David's reinforcer assessments. Top panels depict the assessment of leisure items and bottom panels depict the assessment of edibles. In the left panels, each data point depicts the total number of correct responses in a specific session. The open circles are baseline sessions. The closed squares depict reinforcement sessions with the stimulus ranked first in the MSWO. The open triangles depict reinforcement sessions with the stimulus ranked last in the MSWO. The right panel depicts the break point analyses. The black bar depicts the mean break point (last schedule value completed) across the last three sessions for the stimulus ranked first in the MSWO. The white bar depicts the mean break point across the last three sessions for the stimulus ranked last in the MSWO. The error bars depict the range of break points for each stimulus. SRR = Starting response requirement. PR = progressive ratio step size.

In the assessment of edibles (bottom panels), he consistently completed 17 responses in baseline sessions (M = 17). The number of responses completed per session increased when the stimulus ranked first was used as a reinforcer (*M* = 63.25) but decreased following the initial reinforcement session when the stimulus ranked last was used as a reinforcer (*M* = 30.25). The break point analysis (bottom right panel) indicates that the stimulus David selected first in the MSWO had a higher break point (*M* = 15.33), than the stimulus he selected last (*M* = 9.33), which suggests that the MSWO correctly identified the highest preferred leisure item from the array.

Table 3

Days to Complete Each Reinforcer Assessment

	Reinforcer Assessment		Days to Complete	
Participant	Stimuli	Number of Sessions	From MSWO[a]	From Baseline[b]
Jackson	Edibles	13	22	22
	Leisure	12	12	11
David	Leisure	18	65	49
	Edibles	11	14	11
Connor	Edibles	14	25	24
	Leisure	10	12	9
Laura	Leisure	15	24	23
	Edibles	14	13	9

Note. [a] Days to complete assessment from first MSWO session to last PR session. [b] Days to complete assessment from first baseline session to last PR session.

Results of Connor's reinforcer assessments are depicted in Figure 3. He completed the assessment with edibles first (top panels), and then with leisure items (bottom panels). In the first assessment (top left panel), task completion was initially high in baseline, but dropped in the second and third session (*M* = 48). Responding increased during reinforcement, with a higher number of responses completed per session, on average, when the stimulus ranked first was used as a reinforcer (*M* = 152.8), than when the stimulus ranked last was used as a reinforcer (*M* = 80.25). The break point analysis (top right panel) indicates that the stimulus Connor selected first in the MSWO had a slightly higher break point (*M* = 28.33), than the stimulus he selected last (*M* = 26.67), which suggests that the MSWO correctly identified the highest preferred edible from the array. In the assessment of leisure items (bottom panels), he completed very few responses in baseline (*M* = 18.33). The number of responses completed per session increased when the stimulus ranked first was used as a reinforcer (*M* = 83) but decreased following the initial reinforcement session when the stimulus ranked last was used as a reinforcer (*M* = 18.5). The break point analysis (bottom right panel) indicates that the stimulus Connor selected first in the MSWO had a higher break point (*M* = 25), than the stimulus he selected last (*M* = 5), which suggests that the MSWO correctly identified the highest preferred leisure item from the array.

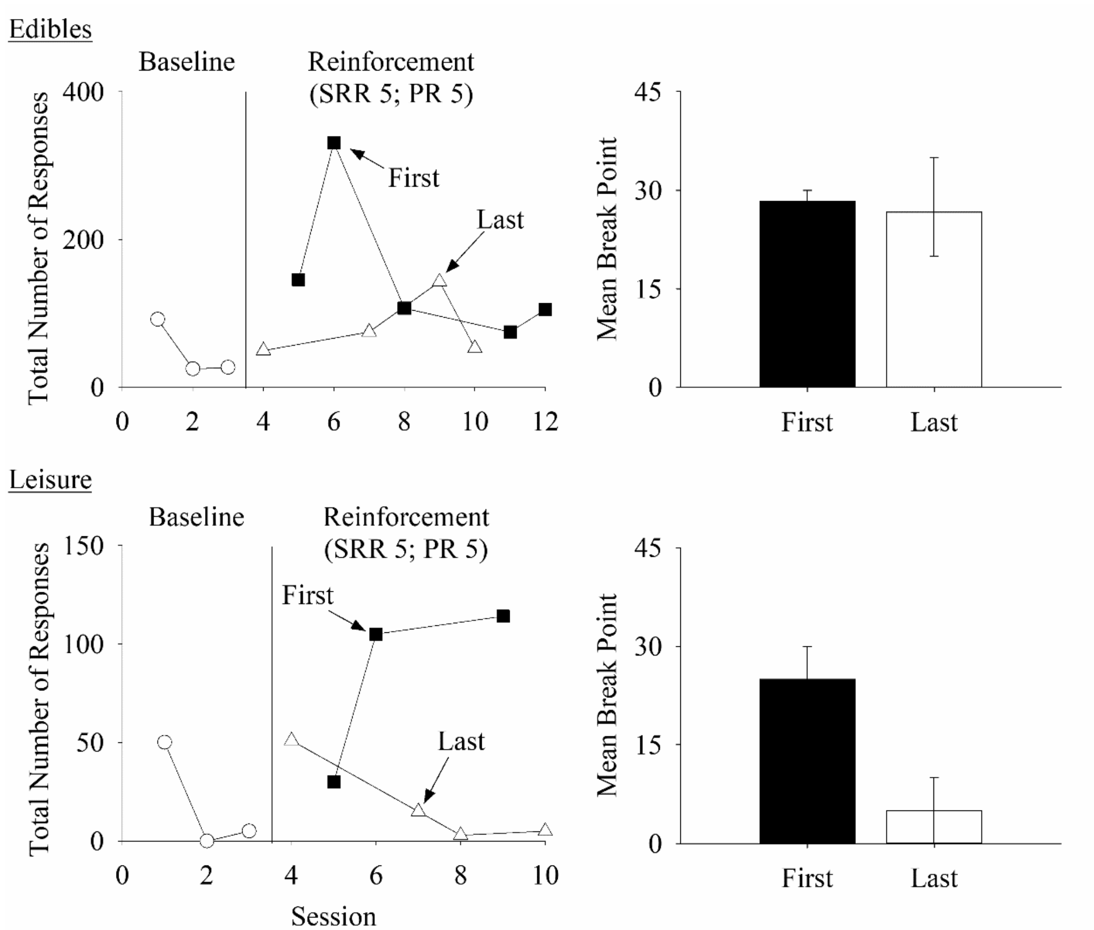

Figure 3. Results of Connor's reinforcer assessments. Top panels depict the assessment of edibles and bottom panels depict the assessment of leisure items. In the left panels, each data point depicts the total number of correct responses in a specific session. The open circles are baseline sessions. The closed squares depict reinforcement sessions with the stimulus ranked first in the MSWO. The open triangles depict reinforcement sessions with the stimulus ranked last in the MSWO. The right panel depicts the break point analyses. The black bar depicts the mean break point (last schedule value completed) across the last three sessions for the stimulus ranked first in the MSWO. The white bar depicts the mean break point across the last three sessions for the stimulus ranked last in the MSWO. The error bars depict the range of break points for each stimulus. SRR = starting response requirement. PR = progressive ratio step size.

Results of Laura's reinforcer assessments are depicted in Figure 4. She completed the assessment with leisure items first (top panels), and then with edibles (bottom panels). In the first assessment (top left panel), the task was completing simple addition. She completed very few responses in baseline (*M* = 1.66). The number of responses completed increased in the first reinforcement session with each stimulus, but then dropped precipitously. After three sessions with each stimulus, the duration of access to the reinforcer was increased from 30 s to 2 min, but responding did not increase in the next session, therefore the PR step size was decreased to 1. Responding remained low when the stimulus ranked first (*M* = 2.33) or the stimulus ranked last (*M* = 1) was used as a reinforcer. Although neither stimulus appeared to function as a reinforcer in this assessment, the break point analysis (top right panel) indicates that the stimulus Laura selected first in the MSWO had a slightly higher break point (*M* = 1.66), than the stimulus she selected last (*M* = 1), which suggests that Laura was saving not the best for last in the MSWO of leisure. Given the limited responding in the first assessment, the task for the reinforcer assessment of edibles (bottom panel) was changed to tracing letters, a reportedly less effortful task. Laura's task completion was initially high in baseline, but eventually dropped (*M* = 56). The number of responses completed per session increased most when the stimulus ranked first was used as a reinforcer (*M* = 145). Laura completed less responses when the stimulus ranked last was used as a reinforcer (*M* = 51). The break point analysis (bottom right panel) indicates that the stimulus Laura selected first in the MSWO had a higher break point (*M* = 35), than the stimulus she selected last (*M* = 23.33), which suggests that the MSWO correctly identified the highest preferred item.

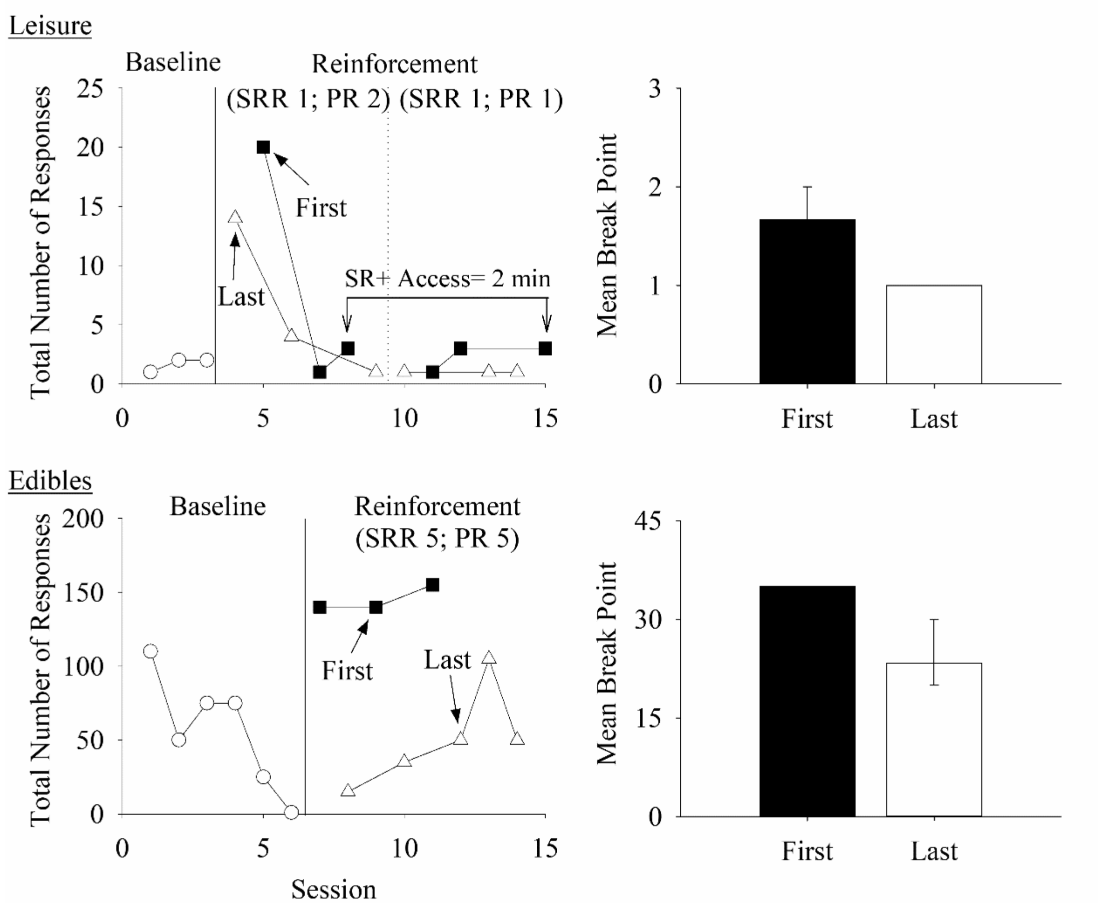

Figure 4. Results of Laura's reinforcer assessments. Top panels depict the assessment of leisure items and bottom panels depict the assessment of edibles. In the left panels, each data point depicts the total number of correct responses in a specific session. The open circles are baseline sessions. The closed squares depict reinforcement sessions with the stimulus ranked first in the MSWO. The open triangles depict reinforcement sessions with the stimulus ranked last in the MSWO. The right panel depicts the break point analyses. The black bar depicts the mean break point (last schedule value completed) across the last three sessions for the stimulus ranked first in the MSWO. The white bar depicts the mean break point across the last three sessions for the stimulus ranked last in the MSWO. The error bars depict the range of break points for each stimulus. SRR = Starting response requirement. PR = progressive ratio step size.

Discussion

In the current study, the experimenters aimed to determine whether four individuals with IDD who have moderate or high verbal ability would save their most preferred item for last in an MSWO preference assessment. Following the MSWO, items selected first and last were assessed under PR schedules to determine reinforcer efficacy.

In our sample, 2 in 4 participants responded more for the leisure item selected last in the MSWO, than for the item selected first. With edibles, 1 in 4 participants responded more for the item selected last, than for the item selected first. These results suggest that at least a quarter of our sample was saving the best for last in the MSWO, even after exhibiting positive time preference in the delay sensitivity assessment.

There were some challenges in interpreting the data from three reinforcer assessments, due to variability (i.e., David's leisure and Connor's edible) or low levels of responding (i.e., Laura's leisure). Toward the end of the assessment, David responded more for the leisure item that he selected last than for that which he selected first. Given the variability in David's responding during the assessment, the degree of overlap between both data paths, and the days to completion of the assessment, however, it is unclear whether he was saving the best for last, or whether the item he selected last gradually became more preferred throughout the assessment. In Connor's case, the difference in mean break points for each edible was not meaningful enough to consider the item selected first as more effective than the item he selected last. We propose two alternative explanations that could account for his data. First, it is possible that Connor liked all the food in the MSWO, and he was simply asked to make difficult decisions regarding the order in which to select them. This is supported by the fact that his responding was variable in the MSWO, and the maximum number of MSWO sessions had to be conducted. Second, it is possible that given the high step size (5), our assessment was not sensitive enough to demonstrate the differences in reinforcer efficacy. With regards to Laura's reinforcer assessment of leisure items, despite the overall low levels of responding, Laura did respond more for the leisure item she selected first, than she did for the one she selected last. Nonetheless, the increase in responding relative to baseline was minimal, for the item that produced an increase (i.e., the item selected first). This is probably due to the difficulty of the task, and the amount of effort it required. Had we used a simpler task, such as the one we used subsequently when we assessed edibles, it is possible that responding would have increased to greater levels.

One of the limitations of the current study is that it involved a small sample of participants. Nevertheless, the prevalence of saving the best for last in this study seems to align with that of prior studies with typically-developing preschoolers and young adults (Castillo, Sun, Frank-Crawford, Rooker, & Borrero, 2020). As such, behavior analysts need to be mindful about the possibility that some individuals will save the best for last in the MSWO, because for such individuals, an MSWO mis-identifies the highest preferred item. Furthermore, this may have implications for identifying effective reinforcers.

Preference assessments are commonly used to identify potential reinforcers for behavioral interventions. In fact, Love, Carr, Almason, and Petursdottir (2009) reported that 65% of early intervention practitioners conduct preference assessments daily, prior to therapy sessions. The MSWO is one preference assessment method that is likely to identify multiple potential reinforcers in minimal time. The MSWO has been recommended over other methods given that it is time efficient, which makes it conducive to frequent implementation, and informative with respect to multiple preferred stimuli. This is particularly useful if practitioners are interested in using a variety of reinforcers within a single intervention to minimize satiation with regard to any single reinforcer (Karsten, Carr, & Lepper, 2011). Because in the MSWO all items are presented simultaneously, one agreed upon requirement is that the individual must be able to scan the entire array before responding (Chazin & Ledford, 2016; Karsten et al., 2011), making this procedure inappropriate for individuals who might impulsively select the first stimulus encountered. Beyond this requirement or, conversely, the “exclusionary criterion,” no further boundaries or limitations have been suggested for the MSWO. No other guidelines have been established to identify appropriate candidates for the MSWO, or ways to interpret responding. If individuals save the best for last during MSWO, however, the results of the preference assessment need to be interpreted with caution. There are some reports of individuals behaving in this manner during MSWOs (Becerra & Fahmie, 2014; Litchmore et al., 2014; Ngur et al., 2018; Pendharkar et al., 2017; Roath & Fritz, 2015; Soldberg et al., 2007), but no systematic evaluations of this phenomenon have been published in academic journals. The current study aimed to address this gap in the literature by comparing the reinforcing efficacy of items selected first and last in an MSWO.

It is important to note that all participants in the current study had intellectual disabilities and relatively moderate verbal ability (i.e., although their standard score on the receptive language measure was moderately low, they spoke in full sentences). Future studies should be conducted to further determine participant characteristics that may predict the likelihood of saving the best for last in an MSWO. For example, future studies could compare the responding of participants with and without intellectual disabilities, with different degrees of verbal abilities (e.g., low, moderate, or high), or with different language measures that may consider both expressive and receptive abilities. Perhaps more systematic evaluations can lead to the development of screening tools that could be used to determine whether an MSWO is appropriate for specific participants, or whether the results need to be interpreted differently.

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82*, 463-496. doi:10.1037/h0076860

Becerra, L., & Fahmie, T. A. (2014, May). *A review and analysis of the consistency of MSWO assessments.* Poster presented at the 40th Annual Convention of the Association for Behavior Analysis International, Chicago, IL.

Bem, S. L. (1967). Verbal self-control: The establishment of effective self-instruction. *Journal of Experimental Psychology, 74*, 485-491. doi: 10.1037/h0024822

Benedick, H., & Dixon, M. R. (2009). Instructional control of self-control in adults with co-morbid developmental disabilities and mental illness. *Journal of Developmental and Physical Disabilities, 21*, 457-471. doi: 10.1007/s10882-009-9166-4

Call, N. A., Trosclair-Lasserre, N. M., Findley, A. J., Reavis, A. R., & Shillingsburg, M. A. (2012). Correspondence between single versus daily preference assessment outcomes and reinforcer efficacy under progressive-ratio schedules. *Journal of Applied Behavior Analysis, 45*, 763–777. doi:10.1901/jaba.2012.45-763

Castillo, M. I, Sun, S., Frank-Crawford, M. A., & Borrero, J. C. (2020). *Save the best for last I: Young adults demonstrate negative time preference - A replication and extension.* Manuscript in preparation.

Castillo, M. I, Sun, S., Frank-Crawford, M. A., Rooker, G. W., & Borrero, J. C. (2020). *Save the best for last II: Whether one saves the best for last depends on outcome category*. Manuscript in preparation.

Chazin, K.T., & Ledford, J.R. (2016). Preference assessments. In *Evidence-based instructional practices for young children with autism and other disabilities*. Retrieved from http://vkc.mc.vanderbilt.edu/ebip/preference-assessments.

Copeland, A. P. (1981). The relevance of subject variables in cognitive self-instructional programs for impulsive children. *Behavior Therapy, 12*, 520-529. doi: 10.1016/S0005-7894(81)80089-5

Cuskelly, M., Gilmore, L., Glenn, S., & Jobling, A. (2016). Delay of gratification: a comparison study of children with Down syndrome, moderate intellectual disability and typical development. *Journal of Intellectual Disability Research, 60*, 865-873. doi: 10.1111/jir.12262

DeLeon, I. G., Frank, M. A., Gregory, M. K., & Allman, M. J. (2009). On the correspondence between preference assessment outcomes and progressive-ratio schedule assessments of stimulus value. *Journal of Applied Behavior Analysis, 42*, 729–733. doi:10.1901/jaba.2009.42-729

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29*, 519-533. doi: 10.1901/jaba.1996.29-519

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*, 1387-1388. doi: 10.1126/science.1151148

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test, fourth edition*. Bloomington, MN: Pearson Assessments.

Francisco, M. T., Borrero, J. C., & Sy, J. R. (2008). Evaluation of absolute and relative reinforcer value using progressive-ratio schedules. *Journal of Applied Behavior Analysis, 41,* 189–202. doi:10.1901/jaba.2008.41-189

Gallagher, T. M. (1999). Interrelationships among children's language, behavior, and emotional problems. *Topics in Language Disorders, 19*, 1-15. doi: 10.1097/00011363-199902000-00003

Gershon, R. C., Slotkin, J., Manly, J. J., Blitz, D. L., Beaumont, J. L., Schnipke, D., … Hirsh-Pasek, K. (2013). IV. NIH toolbox cognition battery (CB): Measuring language (Vocabulary comprehension and reading decoding). *Monographs of the Society for Research in Child Development, 78,* 49-69. doi: 10.1111/mono.12034

Glover, A. C., Roane, H. S., Kadey, H. J., & Grow, L. L. (2008). Preference for reinforcers under progressive- and fixed-ratio schedules: a comparison of single and concurrent arrangements. *Journal of Applied Behavior Analysis, 41,* 163–176. doi:10.1901/jaba.2008.41-163

Harzem, P., Lowe, C. F., & Bagshaw, M. (1978). Verbal control in human operant behavior. *The Psychological Record, 28*, 405-423. doi: 10.1007/BF03394553

Horne, P. J., & Lowe, C. F. (1993). Determinants of human performance on concurrent schedules. *Journal of the Experimental Analysis of Behavior, 59*, 29-60. doi: 10.1901/jeab.1993.59-29

Karsten, A. M., Carr, J. E., & Lepper, T. L. (2011). Description of a practitioner model for identifying preferred stimuli with individuals with autism spectrum disorders. *Behavior Modification, 35*, 347-369. doi: 10.1177/0145445511405184

Keyl-Austin, A. A., Samaha, A. L., Bloom, S. E., & Boyle, M. A. (2012). Effects of preference and reinforcer variation on within-session patterns of responding. *Journal of Applied Behavior Analysis, 45*, 637-641. doi: 10.1901/jaba.2012.45-637

Krasileva, K. E., Sanders, S. J., & Hus Bal, V. (2017). Peabody picture vocabulary test: Proxy for verbal IQ in genetic studies of autism spectrum disorder. *Journal of Autism and Developmental Disorders, 47*, 1073-1085. doi: 10.1007/s10803-017-3030-7

Litchmore, M., Ivy, J. W., & Weaver, M. T. (2014, May). *Reinforcer strength: An examination of assumptions of preference assessment results.* Poster presented at the 40th Annual Convention of the Association for Behavior Analysis International, Chicago, IL.

Loewenstein, G. (1987). Anticipation and the valuation of delayed consumption. *The Economic Journal, 97*, 666-684. doi:10.2307/2232929

Loewenstein, G., & Prelec, D. (1991). Negative time preference. *The American Economic Review, 81*, 347-352.

Loewenstein, G. F., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review, 100*, 91-108.

Love, J. R., Carr, J. E., Almason, S. M., & Petursdottir, A. I. (2008). Early and intensive behavioral intervention for autism: A survey of clinical practices. *Research in Autism Spectrum Disorders, 3*, 421-428. doi:10.1016/j.rasd.2008.08.008

Luria, A. R. (1961). The development of regulatory speech. In J. Tizard (Ed.), *The role of speech in the regulation of normal and abnormal behavior* (pp. 50-96). New York, NY: Pergamon Press.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*, 276-282. doi: 10.11613/BM.2012.031

Meichembaum, D. (1979). Teaching children self-control. In B. B. Lahey & A. E. Kazdin (Eds.), *Advances in Clinical Child Psychology, Vol. 2* (pp. 1-33). doi: 10.1007/978-1-4613-9802-8_1

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science, 244*, 933-938. doi: 10.1126/science.2658056

Ngur, M., Dillon, C. M., & Bowman, L. G. (2018). *Examining unclear preference assessment results due to "saving the best for last" bias.* Poster presented at the 21st Annual Conference of the Maryland Association for Behavior Analysis, Baltimore, MD.

Odum, A. L. (2011). Delay discounting: I'm a k, you're a k. *Journal of the Experimental Analysis of Behavior, 96,* 427-439. doi: 10.1901/jeab.2011.96-423

Olson, M., & Bailey, M. (1981). Positive time preference. *Journal of Political Economy, 89,* 1-25. Retrieved from http://www.jstor.org.proxy-bc.researchport.umd.edu/stable/1837349

Pendharkar, T., Bourret, J. C., Nuzzolilli, A., & Upshaw, S. (2017, May). *Saving the best for last in MSWO preference assessments.* Poster presented at the 43rd Annual Convention of the Association for Behavior Analysis International, Denver, CO.

Petersen, I. T., Bates, J. E., D'Onofrio, B. M., Coyne, C. A., Lansford, J. E., Dodge, K. A., … Van Hulle, C. A. (2013). Language ability predicts the development of behavior problems in children. *Journal of Abnormal Psychology, 122*, 542-557. doi: 10.1037/a0031963

Reed, D. D., & Azulay, R. L. (2011). A Microsoft Excel® 2010 based tool for calculating interobserver agreement. *Behavior Analysis in Practice, 4*, 45–52. doi: 10.1007/BF03391783

Richman, D. M., Barnard-Brak, L., Abby, L., & Grubb, L. (2016). Multiple-stimulus without replacement preference assessment: Reducing the number of sessions to identify preferred stimuli. *Journal of Developmental and Physical Disabilities, 28*, 469-477. doi: 10.1007/s10882-016-9485-1

Roath, C., & Fritz, J. N. (2015, May). *Order of selections in the Multiple-Stimulus Without Replacement preference assessment.* Symposium presented at the 41st Annual Convention of the Association for Behavior Analysis International, San Antonio, TX.

Rodriguez, M. L., Mischel, W., & Shoda, Y. (1989). Cognitive person variables in the delay of gratification of older children at risk. *Journal of Personality and Social Psychology, 57,* 358-367. doi: 10.1037/0022-3514.57.2.358

Siegler, R. S., & Alibali, M. W. (2005). *Children's Thinking.* Englewood Cliffs, NJ: Prentice-Hall.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton Century Crofts.

Soldberg, K. M., Hanley, G. P., Layer, S. A., & Ingvarsson, E. Y. (2007). The effects of reinforcer pairing and fading on preschoolers' snack selections. *Journal of Applied Behavior Analysis, 40,* 633-644. doi: 10.1901/jaba.2007.633–644

Tincani, M. & Travers, J. (2019). Replication research, publication bias, and applied behavior analysis. *Perspectives on Behavior Science, 42,* 59-75. doi: 10.1007/s40614-019-00191-5

Endnotes

[xiii] Psychologists have long theorized that language or verbal behavior, in the form of self-directed speech may help guide behavior to facilitate problem solving (Luria, 1961; Skinner, 1957). Vygotsky viewed the phenomenon of *private speech*, in which children talk aloud to themselves while solving problems or during other activities, as an illustration of children's use of language to regulate their own behavior (Siegler & Alibali, 2005). Further, he suggested that private speech eventually becomes *inner speech*, a silent inner-dialogue, making most thought actually internalized language (Siegler & Alibali, 2005).

[xiv] Establishing specific verbal behavior, such as self-instructions, has been at the center of numerous intervention strategies aimed at improving executive functioning and self-control (Bem, 1967; Copeland, 1981; Diamond, Barnett, Thomas, & Munro, 2007; Meichenbaum, 1979). In a study aimed at promoting the selection of a larger later reward, adults with co-morbid developmental disabilities and mental illness were presented with a written rule that stated, "It is better to pick the [card associated with the larger later reward]" and required to read the rule out loud (Benedick & Dixon, 2009). Self-control training was associated with an increase in the percentage of choices for the larger later reward, and this increase was greatest in the rule condition compared to a no-rule condition (Benedick & Dixon, 2009). Even in the absence of explicit interventions, it is possible that the verbal community (e.g., parents, teachers) reinforce children's spontaneous verbal responses that suggest delaying gratification, such as "saving the best

for last" in the presence of choice opportunities. Subsequently, those verbal responses may serve as self-stated rules that exert control over the individual's behavior.

General Conclusions

This dissertation consisted of a series of studies in which we explored features of sequences that may promote negative time preference (i.e., saving the best for last). In the first study, “Save the Best for Last I: Young Adults Demonstrate Negative Time Preference - A Replication and Extension,” college students responded to hypothetical questions via an online survey. Specifically, we aimed to investigate whether the timing of choices and the number of outcomes in a sequence influence participants’ choice to save the best for last. Part 1 was a replication and extension of procedures described by Loewenstein and Prelec (1991, 1993). Response patterns like those of Loewenstein and Prelec were observed, in that the percentage of participants who saved the best for last decreased when the interval between activities in the sequence increased. There was, however, a significant reduction in the percentage of participants in the current sample who exhibited negative time preference when there were delays involved, compared to those reported by Loewenstein and Prelec.

In Part 2 of the first study, participants were surveyed about their preference for the order in which they would experience hypothetical outcomes with sequences of different sizes (e.g., 3 activities to sequence or 8). As array size increased, the percentage of participants who saved the best for last, or generated a perfectly improving sequence, decreased. Furthermore, there was an interaction effect between the size of the array, and the order of exposure, in that generating perfectly improving sequences for arrays of six, seven, or eight outcomes, was more likely when those where experienced toward the end of the survey. This may be an area worth exploring further, as it may shed light on potential interventions for impulsive behavior. Perhaps, exposing people to sequences of increasing array sizes may promote negative time preference, self-control, or delay of gratification.

Interestingly, the percentage of participants who saved the best for last in Part 2, when there were only two outcomes in the array, was smaller than it was in Part 1 (on all but the second question), which also involved sequences of only two outcomes. Methodological differences may account for these discrepancies. In Part 1, participants were provided a clear description of the different outings in the sequence. In Part 2, however, a list was not provided, and participants were prompted to think of brief activities that they like to varying degrees including favorite activities, things they enjoy but do not love doing, and boring activities they do not enjoy. Therefore, in Part 2, there was variability, within and across participants, in the type of outcomes that they ranked during the survey. It is possible that very different results would have been obtained in Part 2, had a more circumscribed list of outcomes been provided to participants to rank.

Consequently, in the next three studies in the series, described in Paper 2, “Save the Best for Last II: Whether One Saves the Best for Last Depends on Outcome Category,” we decided to study how categorically-different sequences influence saving the best for last, as well as to compare correspondence of responses to hypothetical or real outcomes. Specifically, in Study 1, 192 college students responded to hypothetical questions involving categorically-different lists of outcomes (e.g., noxious stimuli, food, exercise, school work, leisure) via an online survey. A significantly smaller percentage of participants saved the best for last relative to prior studies, including the one described in Paper 1, but the percentage was highest when sequences involved noxious stimuli or food. Next, in Study 2 we examined the correspondence between eight college students’ preference for the order in which they would experience sequences of categorically-different outcomes when those were hypothetical versus real. There was a very strong correspondence in the ranks assigned to the hypothetical and real outcomes, but more variability in the sequences generated. In Study 3 we aimed to determine the order in which four preschoolers would schedule sequences of categorically-different outcomes (i.e., food, leisure, exercise, stories, school work, or a mix of all categories). With academic items, 2 of the 4 participants chose to save the best for last. With leisure items, none of the participants saved the best for last. We found that preschoolers generally interspersed more- and less- preferred activities.

In the series of studies described in Paper 2, roughly a quarter of participants chose to save the best for last. This is quite different from the results of our first study on sequence preferences (Paper 1) in which 64.87% of participants saved the best for last when arranging a sequence of five hypothetical outcomes. Results of these studies suggest that preference for improving sequences, or saving the best for last, may not be as robust as previously indicated, and provide support for prior conclusions that different methods may occasion different response patterns (Andrade & Hackenberg, 2012; Frederick & Loewenstein, 2008).

In Study 2 (Paper 2), half of the participants saved the best for last when scheduling real noxious stimuli, and at least one participant stated in the exit survey that if she were to do it again, she would have scheduled it from least to most preferred. The fact that young adults exhibit negative time preference with noxious outcomes, may be worth noting when scheduling medical interventions within a single session. Findings from prospective or retrospective evaluations of sequences may indicate a general preference for improvement (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993). Consequently, even if the overall pain experienced would not differ, individuals undergoing painful medical treatments may prefer treatments that are ordered in a way such that the pain diminishes over time, rather than gradually increases. Further, their memory of the treatment may be less aversive if relief from the pain is gradual, as would be in the case of an improving sequence, rather than abrupt, as it would be following a worsening sequence (Kahneman et al., 1993). This consideration is important because it is possible that retrospective subjective evaluations of a medical appointment or intervention could influence a person’s morale or even compliance with follow-up appointments and future treatment recommendations. This could all subsequently affect medical outcomes (Redelmeier & Kahneman, 1996). Of course, further extensions are necessary to determine if these things would occur.

One consistent finding across the first four studies in the series is that the percentage of participants who saved the best for last was reliably greater than the percentage of participants who generated perfectly improving sequences. Nevertheless, in the current studies participants ranked and generated sequences just once per category or type of outcome. It is possible that results had differed with repeated measures. Future studies could evaluate (a) response patterns when participants can generate sequences multiple times, (b) retrospective evaluations of sequences, when those are generated by the participants or by the experimenter, and (c) whether experiencing sequences in a preferred order is also associated with changes in performance during the different activities.

Finally, the last study in the series entitled “Save the Best for Last III: Some Children with Intellectual and Developmental Disabilities Save the Best for Last in the MSWO” was related to a more applied problem, involving the identification of preferred items to be used as potential reinforcers in behavior interventions to either teach new skills, or eliminate severe problem behavior. In particular, the purpose of our last study (Paper 3), was to determine whether individuals with intellectual and developmental disabilities who speak in full sentences will save their most preferred item for last when presented with an array of five items and given a chance to access each in a sequence of their own choosing, in a multiple-stimulus without replacement (MSWO; DeLeon & Iwata, 1996) arrangement. In our sample, 1 in 4 participants worked more for the item selected last in the MSWO, than for the item selected first. This suggests that he was saving the best for last in the MSWO, even after exhibiting positive time preference in the delay sensitivity assessment. This was the case for food and toys. Even with a small sample, the prevalence of saving the best for last in the final study seems to align with that of the studies described in Paper 2, with typically-developing preschoolers and young adults. As such, behavior analysts need to be mindful about the possibility that some individuals will save the best for last in the MSWO, because for such individuals, an MSWO mis-identifies the highest preferred item. Furthermore, this may have important implications for identifying effective reinforcers. For individuals who save the best for last, an MSWO may not be the most appropriate arrangement for identifying preferred items. Future studies should be conducted to further determine participant characteristics that may predict the likelihood of saving the best for last in an MSWO. Perhaps more systematic large-scale evaluations can lead to the development of screening tools that could be used to determine whether an MSWO is appropriate for specific participants, or whether the results need to be interpreted differently.

One final concluding point from this series of studies is that saving the best for last does happen. In the two purely hypothetical studies, Paper 1 and Study 2.1, in most cases, at least 1 in 4 participants saved the best for last, in other cases many more. Across all the studies involving real outcomes, at least 1 in 4 participants in all samples was saving the best for last. We believe this is sufficient to be worth further consideration. This dissertation contributes to the existing literature in the field in several ways. In addition to conducting a replication of a seminal study conducted nearly three decades ago, the studies described in this dissertation have provided proof of phenomenon in previously unstudied populations, including preschoolers and individuals with IDD.

References

Andrade, L. F., & Hackenberg, T. D. (2012). Saving the best for last? A cross-species analysis of choices between reinforcer sequences. *Journal of the Experimental Analysis of Behavior, 98*, 45-64. doi: 10.1901/jeab.2012.98-45

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29,* 519-533. doi: 10.1901/jaba.1996.29-519

Frederick, S., & Loewenstein, G. (2008). Conflicting motives in evaluations of sequences. *Journal of Risk and Uncertainty, 37*, 221-235. doi:10.1007/s11166-008-9051-z

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science, 4*, 401-405. doi:10.1111/j.1467-9280.1993.tb00589.x

Loewenstein, G., & Prelec, D. (1991). Negative time preference. *The American Economic Review, 81*, 347-352.

Loewenstein, G. F., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review, 100*, 91-108.

Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, 66, 3-8. doi:10.1016/0304-3959(96)02994-6