DISCOUNTING AND INDEPENDENCE IN PREFERENCES BETWEEN HEALTH SEQUENCES

by

Jonathan R. Treadwell

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

1997

Approved by John M Miyamoto
Chairperson of Supervisory Committee

Program Authorized
to Offer Degree Psychology

Date 6/6/97

UMI Number: 9736390

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

Doctoral Dissertation

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature Jonathan R. Treadwell

Date 6/6/97

University of Washington

Abstract

DISCOUNTING AND INDEPENDENCE IN PREFERENCES BETWEEN HEALTH SEQUENCES

by Jonathan R. Treadwell

Chairperson of the Supervisory Committee
Associate Professor John M. Miyamoto
Department of Psychology

Decision analysts and cost-effectiveness analysts in the health domain often use discounted Quality-Adjusted Life Years (QALYs) as a measure of utility. These analysts commonly assume a positive discount rate: health outcomes in the distant future receive less weight than health outcomes in the near future. In Experiment 1, however, we found that most people actually use negative discounting when choosing between health sequences. Experiment 2 cast doubt on the idea that negative discounting can be explained by desires to maximize savoring and minimize dread. Experiments 3 and 4 gave strong support for the independence assumption of the discounted QALY model, and also replicated the tendency toward negative discounting for health sequences. Taken together, these findings suggest that the general structure of the discounted QALY model is descriptively accurate, but the specific use of a positive discount rate for health sequences should be reconsidered.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

E. Excellent health

MC. Mild coughing

N. No procedure

P. Painful procedure

PC. Persistent coughing

QALYs. Quality-adjusted life years

SH. Severe headaches

ACKNOWLEDGMENTS

The author wishes to thank John M. Miyamoto for numerous helpful discussions about patient preferences and medical decision making. Also, the author expresses his sincere appreciation to Samantha Burns, Charlyne Coke, Tamara Davis, Lynne Hodges, Melissa Moore, and Laura Tiberio for their hard work in helping conduct the four experiments.

DEDICATION

The author wishes to dedicate this dissertation to his mother, Carol Ann Ritter, his father, David Rogers Treadwell Jr., and his brother, David Rogers Treadwell III. Your love, support, and encouragement has helped make this work possible.

CHAPTER 1: INTRODUCTION TO CONCEPTUAL ISSUES

QUALITY-ADJUSTED LIFE YEARS

Consider a patient with stage T3 laryngeal cancer (McNeil, Weichselbaum and Pauker, 1981). There are two possible treatments: surgery or radiation. Unfortunately, it is not immediately clear which treatment should be chosen. If surgery is chosen, the patient has a longer life expectancy but will lose their natural voice. As a result, the patient would have to speak by artificial means such as an electrolarynx (mechanical voice box). On the other hand, if radiation is chosen, the patient will keep their natural voice but will have a shorter life expectancy. Note that this choice involves a tradeoff between how *long* the patient lives and how *well* the patient lives.

A traditional approach to this decision would depend largely (or perhaps completely) on the physician's judgment. The physician might consider many different aspects of the decision, such as how bad it would be to lose one's natural voice, and how important is the difference in life expectancy. These estimates (among others) would be juggled together in the physician's mind to produce a recommendation for treatment.

Recently, health researchers have sought to include the patient more fully in the decision making process. Physicians' estimates may not accurately represent patients' values. Therefore, many researchers have directed questions to the patients themselves to measure critical yet subjective variables such as quality of life. These measurements can then be combined with the physician's expert knowledge of other decisional components (such as the probability of living for X years). Such combinations reflect a belief that health care should be a partnership between physician and patient.

A formal system that incorporates both physician and patient judgments, called "decision analysis", has been developed (McNeil, Weichselbaum, and Pauker, 1981; Weinstein, Fineberg, and Elstein, 1980). This system is based on a theory of rational decision making called Expected Utility Theory (von Neumann and Morganstern, 1947). Each possible outcome of each treatment option is assigned both a probability and a utility. The probability can be taken as the proportion of previous patients who experienced that outcome after receiving that treatment. The utility is a number that is intended to reflect the patient's attitude about that outcome (better outcomes are assigned higher utilities). The analyst computes the "expected utility" of each treatment option, and recommends the option with the highest expected utility.

In medical decision analysis, utility is often measured as Quality-Adjusted Life Years (hereafter abbreviated as "QALYs"). In this model, the number of years for a particular time interval is adjusted by the quality of health during that time interval. For example, 10 years with artificial speech might be "worth" only 9 QALYs because of the decrement in health quality due to the loss of natural speech. The simplest form of the model[1] can be written formally as

[1] One more complicated form of the QALY model was introduced by Pliskin, Shepard, and Weinstein (1980). The equation for that model is: QALYs(living for Y years in health state Q)= b [H(Q)*Y] ^{r} . The "b" is a scaling constant. The parameter r is a risk attitude parameter; the person is risk neutral whenever $r=1$, risk averse whenever $r<1$ and risk prone whenever $r>1$. Risk attitude is not a concept that is central to this dissertation because the data are binary choices between riskless options.

$$QALYs(\text{living for Y years in health state Q}) = Y * H(Q) \qquad \text{Equation 1}$$

where H(Q) is a number between 0 and 1 that represents the perceived quality of life with health state Q. Excellent health is assigned H(Q)=1, and death is assigned H(Q)=0.[2] In the above example, the statement that 10 years with artificial speech is worth 9 QALYs implies that H(artificial speech)=0.9. The QALY measure is a useful and intuitively appealing way to represent both quantity and quality of life with one number. In the framework of decision analysis, a choice dilemma can be resolved by choosing the treatment with the highest expected number of QALYs.

HEALTH SEQUENCES

In the above examples, health state was assumed to be constant over the entire life duration. A more realistic situation is one in which a patient's health changes over time: sometimes the patient is in relatively good health, but other times the patient is in relatively poor health. We use the term "health sequence" to describe this kind of situation. One example of a medical scenario that involves a health sequence is chemotherapy treatment for cancer. The treatment initially causes poor quality of life, and the hope is that eventually the cancer will be cured and health will improve.

The QALY model is easily applied to health sequences. The life duration is divided into intervals; health quality is constant within any interval, but it can change between intervals. QALYs are first computed separately for each interval, and then these QALYs are added together to yield the overall number of QALYs for the entire sequence. The formula is

[2] Many different methods have been used to determine H(Q) for intermediate health states. For a review, see Froberg and Kane (1989).

$$\textit{QALYs}(\text{1st year in health state } Q_1\text{, then } \text{2nd year in health state } Q_2\text{, then } \ldots \text{ } Y\text{th year in health state } Q_Y) = \sum_{i=1}^{Y} H(Q_i) \quad \text{Equation 2}$$

where Y is the duration of the entire sequence, and $H(Q_i)$ is the perceived quality of life in health state Q_i. Each interval lasts for exactly one year, so Equation 2 does not allow health quality to change more frequently than once a year. Obviously, health quality can change more frequently than that; the above equation is simply intended to show the structure of the model as applied to health sequences. For example, consider a health sequence in which the patient lives for 2 years in excellent health, then lives for 3 years in moderate health, and then lives for 1 year in poor health. Assume that H(excellent health)=1, H(moderate health)=0.75 and H(poor health)=0.4. The number of QALYs for this 6-year sequence would be 4.65:

$$\textit{QALYs} = H(Q_1) + H(Q_2) + H(Q_3) + H(Q_4) + H(Q_5) + H(Q_6) = 1 + 1 + 0.75 + 0.75 + 0.75 + 0.4 = 4.65 \quad \text{Equation 3}$$

DISCOUNTING

Many researchers, especially those analyzing cost-effectiveness, use *positively discounted* QALYs as the outcome measure (Weinstein and Stason, 1977). Years in the distant future are given less weight than years in the near future. Because of this differential weighting, the act of positive discounting tends to favor those treatments whose health benefits are realized sooner rather than later. As applied to health sequences, the general formula is

$$\text{Discounted } QALYs(\text{1st year in health state } Q_1, \text{then 2nd year in health state } Q_2, \text{then} \ldots \text{Yth year in health state } Q_Y) = \sum_{i=1}^{Y} \frac{H(Q_i)}{(1+d)^{i-1}}$$ Equation 4

where Y is the sequence duration, $H(Q_i)$ is the perceived quality of life during year i, and d is the discount rate.[3] "Positive" discounting means that d is greater than 0, whereas "negative" discounting means that d is less than 0. Note that in negative discounting, years in the distant future are given *more* weight than years in the near future. Consider again the 6-year sequence in which the first 2 years are lived in excellent health, then the next 3 years are lived in moderate health, and the last 1 year is lived in poor health. If d=0.05 (positive discounting), then the number of QALYs for the sequence would 4.21:

$$\text{Discounted } QALYs = \frac{H(Q_1)}{(1+0.05)^{1-1}} + \frac{H(Q_2)}{(1+0.05)^{2-1}} + \frac{H(Q_3)}{(1+0.05)^{3-1}} + \frac{H(Q_4)}{(1+0.05)^{4-1}} + \frac{H(Q_5)}{(1+0.05)^{5-1}} + \frac{H(Q_6)}{(1+0.05)^{6-1}} = \frac{1}{1.05^0} + \frac{1}{1.05^1} + \frac{0.75}{1.05^2} + \frac{0.75}{1.05^3} + \frac{0.75}{1.05^4} + \frac{0.4}{1.05^5} = 1 + 0.95 + 0.68 + 0.65 + 0.62 + 0.31 = 4.21$$ Equation 5

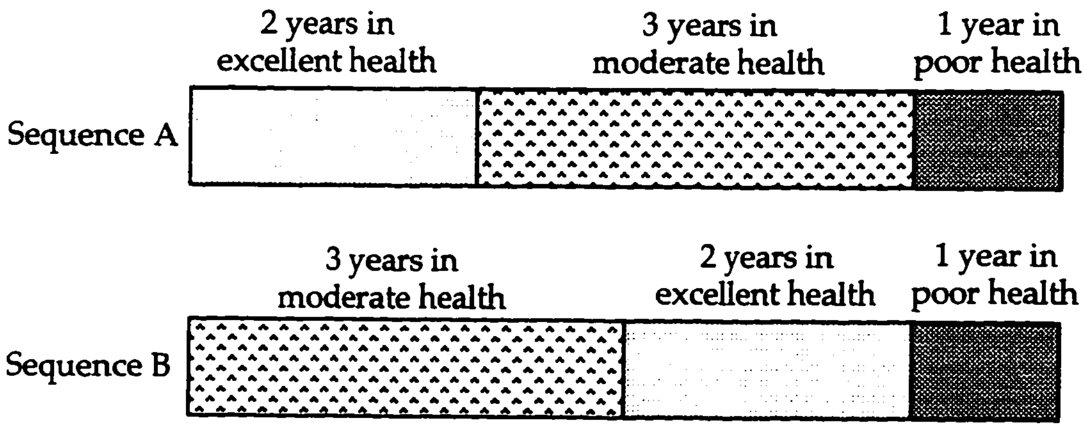
However, if d=-0.05 (negative discounting), then the number of QALYs for the sequence would be 5.19:

[3] The form of Equation 4 implies an exponential discount rate. This is the standard shape of the discount function, but other forms are theoretically possible.

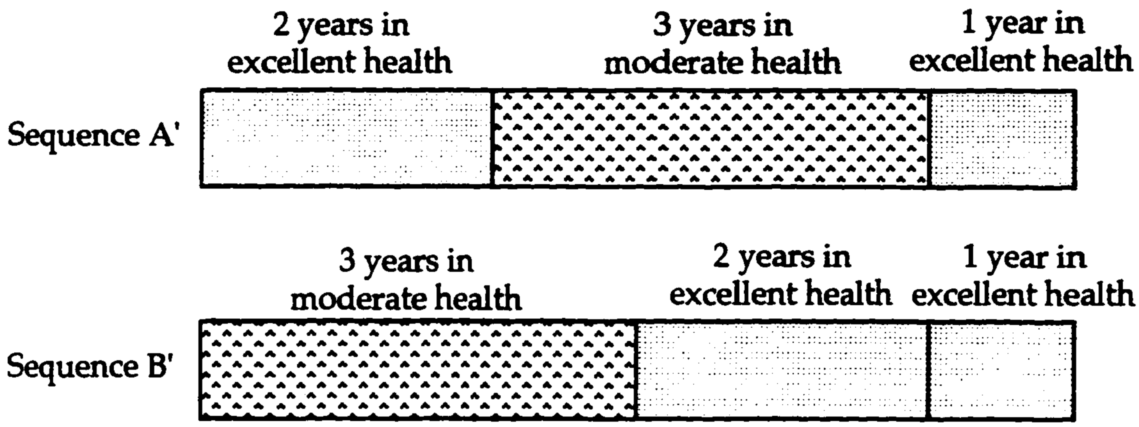
$$\text{Discounted } QALYs = \frac{H(Q_1)}{(1-0.05)^{1-1}} + \frac{H(Q_2)}{(1-0.05)^{2-1}} + \frac{H(Q_3)}{(1-0.05)^{3-1}} + \frac{H(Q_4)}{(1-0.05)^{4-1}} + \frac{H(Q_5)}{(1-0.05)^{5-1}} + \frac{H(Q_6)}{(1-0.05)^{6-1}} = \frac{1}{0.95^0} + \frac{1}{0.95^1} + \frac{0.75}{0.95^2} + \frac{0.75}{0.95^3} + \frac{0.75}{0.95^4} + \frac{0.4}{0.95^5} = 1 + 1.05 + 0.83 + 0.87 + 0.92 + 0.52 = 5.19 \qquad \text{Equation 6}$$

INDEPENDENCE

Regardless of whether positive or negative discounting is used, the QALY model for health sequences implies a condition on preference called "independence." Suppose a patient is faced with a choice between two health sequences, Sequence A and Sequence B:



Notice that the two sequences have the same health state ("poor health") for the last year. Now consider two altered sequences A′ and B′ that have excellent health during the last year:



Satisfying the independence condition means that if one prefers A over B, then one should also prefer A′ over B′. Alternatively, if one prefers B over A, then one should also prefer B′ over A′. Stated generally, independence is the condition that, given two sequences that have the same health state during interval i, preference between them does not depend on the level of health during interval i. The condition is applicable when the tied interval i is at the beginning or the middle of the sequence as well as at the end (as in the example).

Why does the discounted QALY model for health sequences imply independence? Recall that in the model, the discounted QALYs for each interval are computed separately, and then the intervals are added together. Because both sequences A and B end with 1 year in poor health, that last year contributes the same number $\left(\frac{H(\text{poor health})}{(1+d)^5}\right)$ to Sequence A and Sequence B. The last year, therefore, cannot affect the ordering of QALYs(Sequence A) and QALYs(Sequence B). For sequence A′ and B′, the last year adds $\frac{H(\text{excellent health})}{(1+d)^5}$ to both sequences. Again, the ordering of QALYs is unaffected by this addition. The ordering of QALYs is intended to reflect the preference order; hence the QALY model implies that preferences should be insensitive to changes in tied intervals.

The concepts of QALYs, health sequences, discounting, and independence are central to this dissertation. The experiments reported here will be of interest to psychologists, health services researchers, and economists. For psychologists, the study of human preference is an important endeavor because preferences influence behavior. For health services researchers, a critical examination of outcome measures (such as QALYs) helps clarify the usefulness of formal techniques such as decision analysis and cost-effectiveness analysis. For economists, the issues of discounting and independence are implicit in many models of economic behavior. The interdisciplinary nature of these topics suggests that they are fundamentally important.

OVERVIEW OF THE FOUR EXPERIMENTS

Positive discounting of health states is generally believed to be normatively correct. In Experiment 1, we used binary choices between health sequences to determine which kind of discounting subjects tended to use. Surprisingly, the choices of most subjects suggested negative discounting instead of positive discounting.

One possible reason for the prevalence of negative discounting in Experiment 1 involves savoring and dread. Postponing a good health state into the future allows the patient to savor its future occurrence. Also, expediting a poor health state minimizes feelings of dread associated with anticipating a negative event. This explanation, however, requires foreknowledge: that the patient knows in advance about future health states.

In Experiment 2, we manipulated foreknowledge and found that negative discounting occurred even when subjects were told to assume that they would *not* know about future health states in advance.

In Experiment 3, we tested the descriptive validity of the independence condition with health sequences comprised of equal-length intervals (e.g., 10 years with excellent health, then 10 years with mild coughing, then 10 years with severe headaches). In all 27 tests, independence was reliably satisfied.

Experiment 4 investigated health sequences that contained intervals of unequal length. These sequences were constructed to be more sensitive to independence violations. Further, a different medical domain was considered: the short-term treatment of a serious illness. Most of the tests indicated reliable satisfaction of independence, but for 3 tests there were just as many subjects who violated independence as subjects who satisfied independence.

CHAPTER 2: EVIDENCE FOR NEGATIVE DISCOUNTING

INTRODUCTION

How does the timing of an outcome affect its valuation? Is a gain of $100 more appealing now or next year? Is a loss of $100 more disappointing now or next year? For a patient with a chronic disease, is a year in good health worth more now or next year? For a person already in good health, is it better to experience a year of temporary illness now or next year?

Economists of the nineteenth century struggled with similar questions. In an informative historical review, George Loewenstein described the contributions of Rae, Senior, and Jevons to our current understanding of the relation between timing and valuation (Loewenstein, 1992). Rae (1834) recognized the importance of the "desire for accumulation": the extent to which people forego current consumption of resources in favor of assuring future capital. He noted two factors which limit this desire: the brevity and uncertainty of life, and impatience. Life is short; why postpone a reward if it might not eventually be received? Also, people are display impatience because it can be psychologically difficult to delay receipt of a reward. These factors contribute to a feeling that delaying a reward reduces its value.

Rae, however, also recognized two factors that tended to increase desire for accumulation: benevolence and intellectual capacity. An interest in the common good of society rather than personal satisfaction, Rae suggested, increased the desire for accumulation. Also, he thought that a society's degree of intellectual development increased that desire. The basis for these claims is not clear, but the point is this: Rae elucidated some of the conflicted feelings people have about whether to consume now or later.

Senior (1836), like Rae, invoked the notion of impatience in answering the question: why do people pay interest? He saw impatience as an extremely powerful psychological force, and a lender's ability to overcome it should be compensated. The rate of compensation is the interest rate. Jevons (1871) saw that deferral can be a pleasure as well as a pain. The pleasure in deferral comes in anticipation: one can presently enjoy a future reward by imagining and anticipating its occurrence. In this way, the pleasure of immediate consumption is simply replaced by the pleasure of anticipating consumption.

The lesson to be learned from these early economists is that intertemporal choice involves psychological conflict. We are impatient to receive rewards as soon as possible, but we simultaneously see the value in deferring gratification. We would like to postpone some unpleasant experiences, but others are best expedited in order to get them over with. Context plays a large role in determining which impulse wins out. For example, there is no obvious reason to delay receipt of a $100 prize, but there are good reasons to delay the experience of kissing your favorite movie star (e.g., savoring). An incurable illness is best postponed as far into the future as possible, but why not schedule a dental appointment sooner so you can be done with it?

NORMATIVE ISSUES

There is general agreement that money now should be worth more than money in the future because money now can be invested to make more money. For example, given a choice between receiving $100 today or

$100 in one year, most people would prefer the former because the $100 today can be deposited into a savings account to accrue interest for a year. This implies a positive discount rate: one should prefer to expedite monetary gains and postpone monetary losses. Setting the monetary discount rate at the current interest rate after adjusting for inflation (usually about 0.05) accounts for investment opportunity.

Cost-effectiveness analysts typically take a societal perspective: how can we as a society most efficiently allocate scarce resources to health care? Each health program is assigned a ratio of its associated costs to its associated health benefits which is used to measure the program's efficiency (Drummond, Stoddart, and Torrance, 1987; Weinstein and Stason, 1977). Competing programs often differ with respect to the timing of health benefits. For example, the benefits of a cancer treatment program are realized sooner than the benefits of a cancer prevention program. Money spent on any health program has an opportunity cost: it could have been invested otherwise to create more money. Health benefits, therefore, should be discounted according to how far into the future they occur.

There is controversy, however, about what should be the discount rate for life years. Many cost-effectiveness analysts argue that it should be the same as the discount rate for money. Keeler and Cretin (1983) showed that given certain assumptions, using a smaller discount rate for life years than for money results in a paradox in cost-effectiveness analysis. To avoid this paradox, they argued, the two rates should be equal.

Because Keeler and Cretin's (1983) argument is so widely cited, we will describe it in more detail. They made two key assumptions (Keeler and Cretin, 1983, p. 303). First, they assumed that any feasible health program can be delayed indefinitely and it will still be feasible. In other words, after any delay, a program's benefits will still be realized and its costs will still be incurred. Second, they assumed that it is always possible to achieve greater benefits by incurring greater costs. In other words, you can always get more by spending more.

Given these assumptions, they considered the case where health benefits are discounted at a lower rate than costs. In this situation, they showed that any program can be made more cost-effective by delaying it. Thus, if programs were instituted solely on the basis of cost-effectiveness, no program would ever be implemented! This is obviously an unacceptable and absurd result. They assert that the source of the absurdity is the disparity in discount rates.

Many researchers, however, have not accepted the claim that the discount rates must be equal. Roth, Ing, and Ross (1978) stated that the discounting of life years is a questionable value judgment on the part of the decision analyst. Such value judgments, they suggest, have no place in the framework of formal systems (such as decision analysis or cost-effectiveness analysis) that emphasize a decision maker's values over those of an analyst.[4]

Coyle and Tolley (1992) noted that Keeler and Cretin's (1983) proof addresses the issue of *when* to invest money in a health care program. Most evaluations instead involve *whether* money should be invested in any of a set of possible programs. They also questioned the assumption that a program's benefits can be realized at any time in the future. They decided that there is no theoretical basis for claims that the discount rates for money and health should be equal (or unequal) (see also Parsonage and Neuberger, 1992).

---

[4] Using undiscounted life years, it should be recognized, would also be a value judgment on the part of the decision analyst. It seems, therefore, that an analyst's values are an inevitable component of the process.

Consequently, they recommended the use of several discount rates for both costs and benefits via sensitivity analysis.

Ganiats (1994) recognized that allocation decisions are often based on whether a program is *acceptably* efficient rather than whether it is *maximally* efficient. Therefore, even if a program's cost-effectiveness ratio might be smaller (better) if it were delayed, it might still be implemented sooner because it meets an acceptable level of efficiency. Also, Ganiats (1994) critiqued Weinstein and Fineberg's (1980) assertion that equal discount rates are needed in order to maintain a constant relation between money and health over time. Such a constant relation may be unrealistic. If a society's health improves over time while its economy is stable, then the relation between money and health is nonconstant. In that situation, perhaps the discount rate for health should be lower than the rate for money.

Williams (1981), Goodin (1982), and Cairns (1992) stressed the importance of the *tradability* of money and health. To what extent can we improve health as a direct result of spending money? If these two goods are perfectly tradable, then there is good justification for requiring that their discount rates be equal. Cairns (1992) hypothesized, however, that trading opportunities are rather scarce. To the extent that this is true, the discount rates should not be required to coincide.

This controversy might be resolved by considering the issue of perspective (Redelmeier, Heller, and Weinstein, 1994). From a societal perspective, money and health are relatively tradable, and so there is good justification for equating their discount rates. From an individual perspective, however, it is rarely possible to trade directly money for health (e.g., one cannot simply pay a physician to cure an incurable illness).

Therefore, there should be room for a health discount rate that is not necessarily equal to the monetary discount rate.

DESCRIPTIVE ISSUES

Regardless of how the timing of health outcomes *should* affect their valuations, there is still the issue of how those valuations *actually are* affected. Decision analyses and cost-effectiveness analyses are intended to represent decision makers' values. Therefore, there should be some correspondence between the models' assumptions about time preference and the descriptive realities of time preference. Some previous researchers have investigated the relation between time preference and reported health behaviors. Others have actually measured subjects' discount rates for health based on responses to hypothetical questions.

It is possible to infer a subject's discount rate for either money or health based on that subject's expressed preferences. In the money domain, the investigator determines the amount of money $Y to be received after a delay of T years that is preferentially equivalent to receiving a specific amount of money $X now. In the health domain, the investigator determines the duration of symptom relief Y to begin after a delay of T years that is preferentially equivalent to a specific relief duration X to begin now. In either case, the inferred discount rate is

$$d = \left(\frac{Y}{X}\right)^{1/T} - 1 \qquad \text{Equation 7}$$

Note that d>0 if Y>X, and d<0 if Y<X. This discount rate d is exactly the same discount rate d that was introduced in Equation 4 (page 5). If a subject is indifferent between receiving $110 in two years and receiving $100 now, then

the inferred discount rate is $d_{Money} = \left(\frac{110}{100}\right)^{\frac{1}{2}} - 1 = 0.049$. If a subject is indifferent between receiving 6 months of symptom relief starting in two years and receiving 5 months of symptom relief starting now, then the inferred discount rate is $d_{Health} = \left(\frac{6}{5}\right)^{\frac{1}{2}} - 1 = 0.095$.[5, 6]

Fuchs (1982) pointed out that many health behaviors (such as exercising, dieting, and going to the dentist) involve the acceptance of discomfort today in exchange for future health benefits. Time preference, therefore, might help explain variability in these kinds of behaviors. He first measured subjects' discount rates for money as revealed by their choices between [$X now] and [$Y in T years] where X<Y. Then he obtained self-reports of the frequency of various health behaviors (e.g., "When did you have your last dental checkup? 1) Within the last year, 2) About one or two years ago, 3) About three to five years ago, 4) More than five years ago). He hypothesized that subjects who tended to choose money now over more money later would less often engage in future-oriented health behaviors. Some correlations were statistically reliable in the predicted direction (N=508), but the proportion of explained variance in health behavior was consistently tiny (no R^2 was greater than 0.05). It seems that monetary discount rates bear only a small relation to these kinds of health behaviors.

---

[5] It makes no difference that the "6" and "5" are on the scale of months but the "2" is on the scale of years. Because Y is divided by X, the same result would be obtained if "6 months" were re-expressed as "0.5 years" and "5 months" were re-expressed as "0.417 years".

[6] Because the health outcomes are received over time, they should be discounted. The formula does accurately describe how health discount rates have been computed by previous researchers.

Christensen-Szalanski (1984) found that a patient's stated values for a future outcome can be quite different than their stated value for that outcome when it is actually occurring. Pregnant women were asked to rate two things on a 0-100 scale: the importance of avoiding anesthesia during labor, and their concern about avoiding pain during labor. These rating were given one month before labor, during labor, and one month after labor. Ratings both before and after labor favored the avoidance of anesthesia, but during labor the preference pattern switched in favor of avoiding pain. The switch was larger for first-time mothers than for experienced mothers. These results highlight the sensitivity of preferences to the timing of the preference assessment.

Two studies (Rose and Weeks, 1988; Chapman and Elstein, 1995) measured both monetary and health discount rates in the *gain* domain. In these studies, a "health gain" meant temporary relief from a chronic medical problem. For example, subjects in the Rose and Weeks (1988) study were told to imagine receiving 10 weeks of relief from chronic back pain. Each subject's discount rate was measured by determining how many weeks of relief to be received in the future was preferentially equivalent to receiving 10 weeks of relief starting immediately. If more than 10 weeks were required to offset the delay, then the subject was inferred to have a positive discount rate. Conversely, if fewer than 10 weeks were needed to produce indifference, then the discount rate was inferred to be negative. They found generally high positive discount rates for both money and health. Half of the subjects had monetary discount rates greater than 0.40, and 75% of the subjects had health discount rates greater than 0.40. Only a small percentage of subjects (5% for money and 8% for health) had negative discount rates.

Chapman and Elstein (1995) told subjects to imagine they had been in a poor health state for the past two years.[7] Two treatments could each restore them to full health for some limited amount of time. Treatment A's benefit would occur immediately, whereas Treatment B's benefit would occur after a specific delay. Given both the duration of Treatment A's benefit (e.g., 6 months in full health) and the delay associated with Treatment B (e.g., 1 year from now), the subject was told to state what duration of benefit for Treatment B would make the two treatments equally preferable. As in the Rose and Weeks (1988) study, both monetary and health discount rates in the domain of gains were strongly positive. The geometric mean of health discount rates was a surprisingly high 1.24 (see Chapman and Elstein, 1995, p. 378), indicating that Treatment B's benefit generally needed to be quite large in order to compensate for its delay. Monetary discount rates were positive but less extreme.

What about time preference for health *losses*? Three studies (Lipscomb, 1989; Cairns, 1992; Redelmeier and Heller, 1993) concentrated their efforts on health losses. Lipscomb (1989) instructed healthy subjects to rate unpleasant health scenarios on a 0-10 scale. In each scenario, an initial interval in excellent health was followed by an interval in poor health (defined precisely). Because subjects did not actually provide indifference points but instead rated each scenario individually, Lipscomb (1989) could not compute discount rates. However, if a subject's ratings improved as the length of the initial interval increased, positive time preference would be suggested. Indeed, that was the dominant pattern in the data. Lipscomb (1989) suggested that one could measure "pure time preference" by regressing

---

[7] The poor health state was described in detail. It involved restricted diet, frequent medications, occasional insomnia, and other general health problems (Chapman and Elstein, 1995, p. 386).

the ratings onto a complete set of uncorrelated predictor variables and examining the size of the slope of the "initial interval" predictor variable.

Unlike Lipscomb (1989) who used a rating task, Cairns (1992) used a preference matching task and, therefore, could compute discount rates for health losses. First, he described to each subject a poor hypothetical health state. Then the subject was asked to consider two scenarios, A and B. In Scenario A, they would first live in excellent health for 20 years, then would live in the poor health condition for Y days, then would live in excellent health for 22 years, then would die. In Scenario B, the subject would first have only 2 years in excellent health, then 90 days in the poor health state, then 40 years in excellent health, then would die. The subject was asked to provide the value of x for which he/she was indifferent between the two scenarios. He then took the discount rate to be $d_{Health} = \left(\frac{Y}{90}\right)^{1/18} - 1$ (18 is the difference in time delay). Health discount rates were not reliably different from 0. In a related monetary task, however, subjects had reliably positive discount rates. Unlike gains, for which the discount rate was greater for health than for money (Rose and Weeks, 1988; Chapman and Elstein, 1995) for losses the discount rate was greater for money than health (Cairns, 1992).

Redelmeier and Heller (1993) presented subjects with three negative yet temporary health outcomes: colostomy, blindness, and depression. They computed discount rates for each of two preference assessment methods: categorical ratings and standard gamble.[8] They found that 62% of the discount rates equaled 0, 28% were greater than 0, and 10% were less than 0. Surprisingly, only 12% of the discount rates fell in the

[8] As stated above and also by Weinstein (1993), it is not appropriate to infer a discount rate from rating scale data.

interval between 0 and 0.10. Many cost-effectiveness analysts only consider discount rates in this range. Based on their findings, Redelmeier and Heller (1993) strongly questioned the common assumption of small and positive exponential discount rates in cost-effectiveness analyses.

Two studies (MacKeigan, Larson, Draugalis, Bootman, and Burns, 1993; Chapman, 1996a) have directly compared time preference for health gains to time preference for health losses. MacKeigan et al (1993) used rating scale data to suggest that 1) time preference is more strongly positive for health gains than for health losses, and 2) for a brief health loss (e.g., 2 days in a poor health state), *negative* time preference is observed. That is, subjects preferred to have the brief health loss sooner rather than later. Chapman (1996a) used a matching task to infer health discount rates for both losses and gains, and obtained similar results.

SEQUENCES

In all of the aforementioned studies, a single interval of health was delayed. Researchers inferred the direction and magnitude of time preference by determining what interval duration was required to compensate for the delay. It is also possible to investigate time preference using a *sequence* of health states instead of just a single delayed interval.

Suppose an individual's valuation of a health sequence is very sensitive to what happens at the beginning of the sequence but insensitive to what happens at the end. The initial intervals, therefore, receive more weight than later intervals. This kind of valuation suggests positive time preference: later intervals receive less weight because they are heavily discounted. On the other hand, if an individual's sequence valuation is more sensitive to later intervals than to early intervals, negative time preference would be suggested. Such an individual would prefer to postpone gains because then the gains would receive more weight, and also would prefer to expedite losses because then the losses would receive less weight.

Evaluations of sequences generally suggest negative time preference (Kahneman, Fredrickson, Schreiber, and Redelmeier, 1993; Loewenstein, 1987; Loewenstein and Prelec, 1991; Loewenstein and Prelec, 1993; Loewenstein and Sicherman, 1991; Redelmeier and Kahneman, 1996; Ross and Simonson, 1991; Varey and Kahneman, 1992). In a particularly intriguing experiment, Kahneman et al (1993) had subjects immerse a hand into very cold water (14 degrees C., which is painful) for one minute. In a second trial (order was counterbalanced), subjects immersed the other hand in the same water for one minute, and then kept that hand in the water for thirty seconds longer. During the last thirty seconds, the water temperature was gradually raised to a perceptibly warmer yet still painful temperature of 15 degrees C.

When asked which painful trial they would prefer to repeat, 22 of 32 subjects (69%) actually chose to repeat the *longer* trial even though that trial had a longer pain duration. For those subjects, the gradual temperature increase at the end of that trial seemed more important than the fact that the final thirty seconds were still painful. Therefore, they seemed to give more consideration to the end of a sequence and less consideration to its overall duration. Redelmeier and Kahneman (1996) found similar results with patients undergoing either colonoscopy or lithotripsy.

Chapman (1996b) investigated health sequences with much longer durations. Each sequence was divided into 12 5-year intervals to begin at age 20 and end at age 80. She presented subjects with 16 distinct sequences and instructed them to rate the attractiveness of each sequence on a scale from 0 to 100. The level of health during any interval was displayed as a vertical bar whose height ranged from 1 ("very poor health, just barely better than death") to 10 ("perfect health, the best you could imagine"). The sum of the bar heights in each sequence was 66, so the sequences were equal with respect to total quality of health. Some of the sequences were "decreasing" (e.g., health quality ratings of 10, 9.1, 8.3, 7.5, 6.7, 5.9, 5.1, 4.3, 3.5, 2.7, 1.9, 1), whereas others were "increasing" (e.g., health quality ratings of 1, 1.9, 2.7, 3.5, 4.3, 5.1, 5.9, 6.7, 7.5, 8.3, 9.1, 10). Subjects tended to rate the decreasing health sequences (mean rating 58) *higher* than the increasing health sequences (mean rating 34). That pattern of results suggests positive time preference.

In a second experiment, however, Chapman (1996b) examined sequences whose duration was 1 year (divided into 12 one-month intervals) instead of 60 years. Subjects were told to imagine they needed to be treated for a medical condition over the upcoming year, and that their health would return to normal at the end of the year. Ratings for the decreasing one-year sequences (mean rating 45) were *lower* than those for the increasing one-year sequences (mean rating 64). That pattern suggests negative time preference. Chapman (1996b) explained these results via expectations: subjects tended to give higher ratings to sequences that matched their expectations about future health. Lifetime health profiles are usually deteriorating, and a decreasing health sequence matches that expectation. For a one-year sequence, however, subjects expect their health to gradually improve as the hypothetical treatment takes its effect, and an increasing sequence matches that expectation.

There are two important aspects of Chapman's (1996b) method that should be recognized. First, the levels of health quality in each sequence were abstract. A health quality of "5" on a 1-10 scale has little inherent meaning. Second, subjects did not make direct choices between sequences, but rather they rated sequences individually. Of course, preferences might be inferred from the ordering of ratings. It has been shown, however, that ratings can be inconsistent with choices (Goldstein and Einhorn, 1987; Tversky, Slovic, and Kahneman, 1990).

In Experiment 1, we investigated time preference as revealed by direct choices between health sequences with well-defined levels of health quality. We used a 30-year time span starting at age 20 to minimize lifetime health expectation effects. In other words, we suspect that most 20-year olds do not expect their health to deteriorate very much until after age 50.

We did not compute discount rates in Experiment 1 because we did not measure indifference points. Suppose, however, that the discounted QALY model (Equation 4) is an accurate representation of how subjects evaluate sequences. Each subject assigns a unique H(Q) to each of the health states, and also each subjects uses a unique discount rate d. Many of the sequence choices in Experiment 1 (26 of the 30 choices) can be used to infer whether a subject implicitly used a positive discount rate or a negative discount rate. Therefore, we inferred the *valence* of discount rates but did not attempt to estimate the *magnitude* of those rates.

METHOD

SUBJECTS

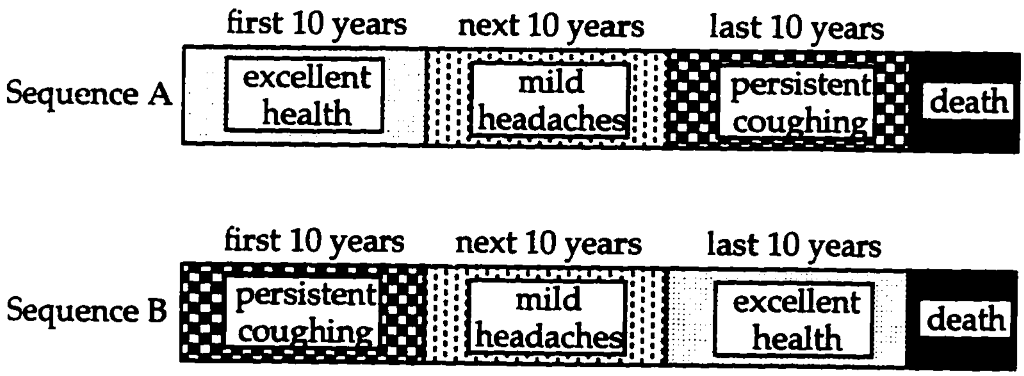
46 University of Washington students participated. All were enrolled in Psychology courses and received extra credit for their participation.

Health States

Three health states were used: excellent health, mild headaches, and persistent coughing. Headaches and coughing were chosen because it is reasonable to expect that all participants had some experience with these conditions. The health state descriptions are reproduced in Appendix A.

Design and Materials

Each health sequence lasted for 30 years, and the 30 years were divided into three ten-year intervals. In each sequence, there was a different health state for each ten-year interval. Because only three health states were used, this produced six distinct health sequences. All 15 possible pairs of the six sequences ("6 choose 2"=15) were presented one at a time to each subject (the "first set"). Then, these 15 pairs were all presented a second time (the "replicated set") to investigate the consistency of subjects' choices. The pair below illustrates the manner in which sequences were presented to subjects:



As shown, the sequence in the "A" slot was located above the sequence in the "B" slot. For purposes of counterbalancing, two sets of stimuli were created: "AB" and "BA". Half of the subjects received the AB set, and the other half received the BA set. The assignment of sequences to the slots was opposite for the two stimulus sets. For example, the AB set contained a pair of sequences in which the A sequence was "Excellent health, Mild headaches, Persistent Coughing" and the B sequence was "Excellent health, Persistent Coughing, Mild Headaches". In the BA set, the former sequence occupied the "B" slot and the latter sequence occupied the "A" slot. This counterbalancing was used to control for any subject bias in choosing sequences based solely on spatial location.

Each subject saw a different random order of the first 15 pairs. The order was randomized anew for the replicated set of choices. To ensure that at least ten choices intervened between the two presentations of any particular pair, the second randomization was performed separately for three blocks of five pairs per block. In other words, the first five pairs of the first set were placed in a new randomized order and were presented as the first five pairs in the replicated set, and the same procedure was used for the next five pairs and for the last five pairs.

Procedure

Subjects were told to imagine having an incurable illness in which they would have only 30 years of remaining life. Subjects read brief descriptions of each health state, and were then asked which of the two intermediate health states (persistent coughing or mild headaches) would be worse. It was emphasized in the instructions that only one of the three health states would be experienced in any ten-year interval. For each of 30 pairs of sequences (fifteen for the first set and fifteen for the replicated set), subjects were asked which sequence would be preferred if they actually had the illness. After all 30 choices had been made, subjects were debriefed and given credit for participating.

RESULTS

Subjects did not completely agree about which of the two intermediate health states would be worse. Whereas 31 subjects (67%) stated that persistent coughing would be a worse state of health than mild headaches, the other 15 subjects (33%) gave the reverse ordering. For the first group, persistent coughing was reclassified as the "poor" health state and mild headaches was reclassified as the "moderate" health state. In contrast, for the second group, persistent coughing was reclassified as the "moderate" health state and mild headaches was reclassified as the "poor" health state. These general labels permitted an overall data analysis of all 46 subjects for 15 pairs of sequences.[9] For example, the labels permitted computation of the overall proportion of subjects who chose the sequence "excellent, moderate, poor" over the sequence "poor, moderate, excellent" even though "moderate" (and "poor") meant different things for different subjects.

Figure 1 displays the overall analysis for each of the 15 pairs. The sequences along the top (and the left side) are placed in order according to a lexicographic rule (Coombs, Dawes and Tversky, 1970). In this rule, two sequences are compared on the third interval; the sequence with the better health state during that interval is placed ahead of the other. If the sequences are tied on that third interval, then the second interval is used to determine the ordering.

One advantage of constructing the figure in this manner is that, for 13 of the 15 cells, *positive discounting implies that the row sequence should be preferred over the column sequence.* In contrast, for the same 13 of

[9] One subject (in the group of 31) skipped the choice of "excellent, moderate, poor" versus "excellent, poor, moderate". Therefore, for that choice, there were only 45 subjects.

the 15 pairs, negative discounting implies that the column sequence should be preferred over the row sequence. These 13 pairs will be referred to as "diagnostic" pairs because the two hypotheses make opposite predictions for them. An example proof of one of these implications is given in Appendix B. Asterisks in the figure indicate the two pairs for which these implications do not necessarily hold. Appendix C describes a counterexample for one of these two pairs.

Each cell in Figure 1 contains a pie chart that displays three proportions: the proportion of subjects who chose the column sequence over the row sequence *twice* (indicated by the striped area of the pie), the proportion of subjects who chose the row sequence over the column sequence *twice* (indicated by the black area of the pie), and the proportion of subjects who did not replicate their choice (indicated by the white area of the pie). Note that *positive discounting implies that the pies should be mostly black, whereas negative discounting implies that the pies should be mostly striped.*

The data indicate that more subjects tended to base their choices on negative discounting. For example, the pie on the bottom left corresponds to the choice between a fully decreasing sequence (the favored option according to positive discounting) and a fully increasing sequence (the favored option according to negative discounting). 26 subjects (57%) chose the increasing sequence twice, 16 subjects (35%) chose the decreasing sequence twice, and 4 subjects (9%) did not replicate their choice. For 12 of the 13 diagnostic pairs, the proportion of subjects who repeatedly chose the negative discounting option was higher than the proportion of subjects who repeatedly chose the positive discounting option. For the two nondiagnostic pairs, subjects tended to choose sequences that gave excellent health in either interval 1 or interval 3 over sequences that gave excellent health in interval 2.

One way to consider the data in Figure 1 is that perhaps some subjects consistently contributed to the striped area of each pie, whereas other subjects consistently contributed to the black area of each pie. In other words, it may be that one group of subjects consistently chose negative discounting options, whereas a separate group of subjects consistently chose positive discounting options. To investigate this possibility, we computed a "discounting score" for each subject. This score should not be confused with a discount rate, which is a different measure. For each of the 26 choices associated with the 13 diagnostic pairs, we counted the number of times that a subject chose the positive discounting option. Then we subtracted 13 from that total to yield that subject's discounting score. The score ranged from -13 (indicating maximum negative discounting) to 13 (indicating maximum positive discounting). A score of 0 indicated no systematic preference for either positive or negative discounting.

A frequency distribution of the discounting score appears in Figure 2. The distribution is impressively bimodal: there was a negative discounting group (N=30) and a positive discounting group (N=16). This is strong evidence for the idea that subjects tended to adopt one of these two strategies. Nine subjects chose the negative discounting option for all 26 diagnostic pairs. A 95% confidence interval around the observed percentage of negative discounting subjects (65%) had a lower bound of 51% and an upper bound of 79%.

Figures 3 and 4 display the pie charts for the two subject groups separately. Note that, for the negative discounting group (Figure 3), all 13 of the pies corresponding to diagnostic pairs are mostly striped. In contrast, for the positive discounting group (Figure 4), those same 13 pies are all mostly black. Of course, these figures are not surprising because subjects were split into two groups based on the discounting score. They are useful, however, to display the striking difference between the groups' preference patterns.

DISCUSSION

The question, then, regarding whether or not people generally use positive discounting when choosing between health sequences cannot be answered with a simple yes/no answer. Some people consistently do, but other people consistently use negative discounting. These results suggest that more people discount negatively.

Recall the suggestion in the Introduction that intertemporal choice involves conflict. One impulse compels us to postpone an unpleasant experience, but another impulse compels us to expedite that experience. It seems that for subjects in Experiment 1, the resolution of that conflict was not a compromise but rather was a simple strategy choice. Subjects did not tend to alternate between choosing positive discounting options and choosing negative discounting options. If they had, the frequency distribution in Figure 2 would have been bell-shaped, but instead it was bimodal. For any particular subject, one of the two strategies consistently prevailed. There may be personality characteristics that underlie this strategy choice. Such variables were not measured, however, and so it is not possible to determine their influence.

Lastly, we note two aspects of the choice situation in Experiment 1 that probably increased the incidence of negative discounting. First, choice were made in the domain of losses: subjects were healthy undergraduates who considered sequences of mediocre health quality. Previous research has shown that losses generally produce lower discount rates than gains (MacKeigan et al, 1993; Chapman, 1996a). Second, each option was a sequence instead of a single interval of health. Other research has found lower discount rates for sequences than single outcomes (Loewenstein and Prelec, 1991).

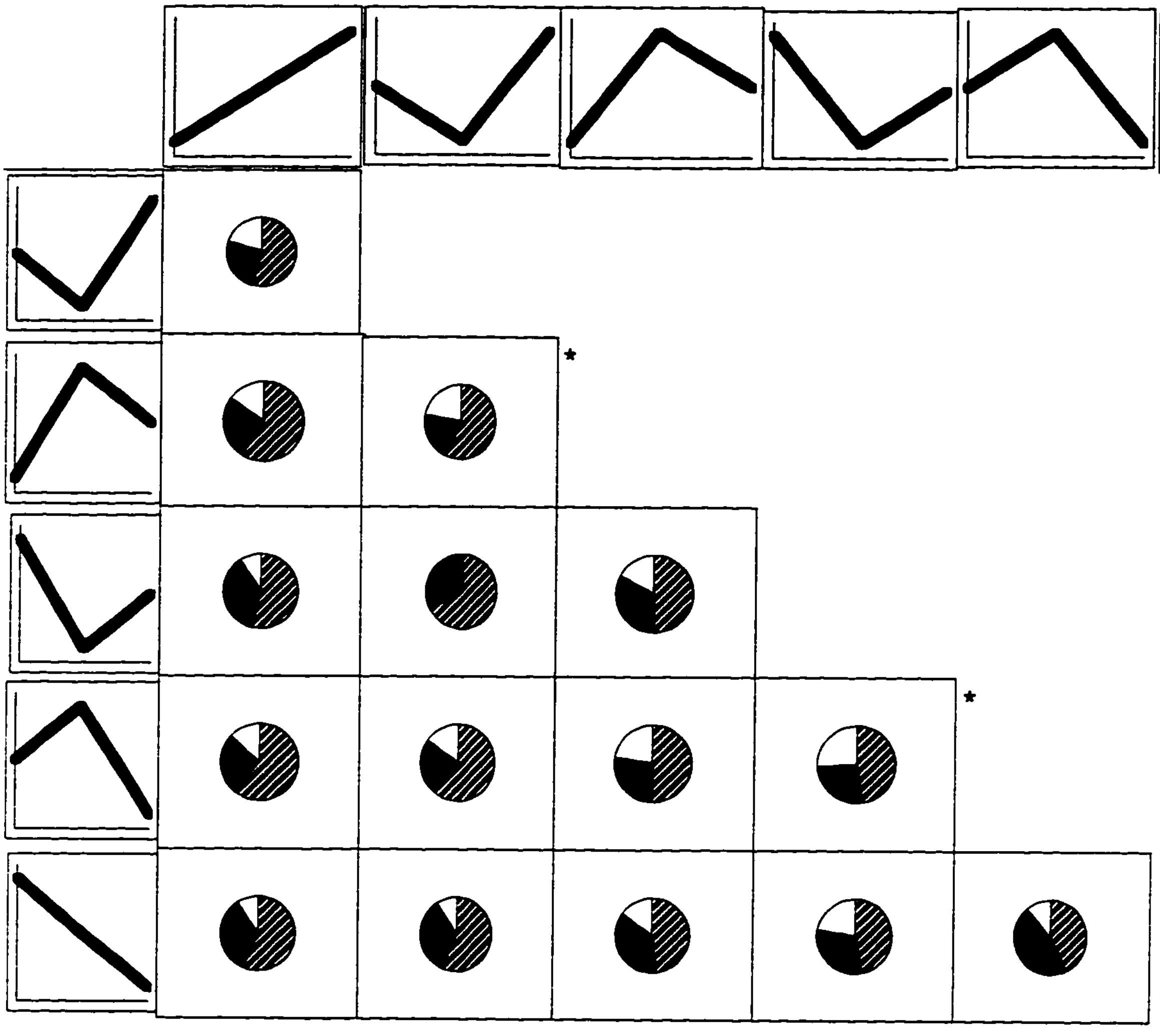


Figure 1: Preference Patterns of All Subjects in Experiment 1

▨ proportion of subjects who chose the column sequence over the row sequence twice

■ proportion of subjects who chose the row sequence over the column sequence twice

□ proportion of subjects who did not replicate choice

* For the pairs marked with an asterisk, neither positive nor negative discounting necessarily imply which sequence should be preferred.

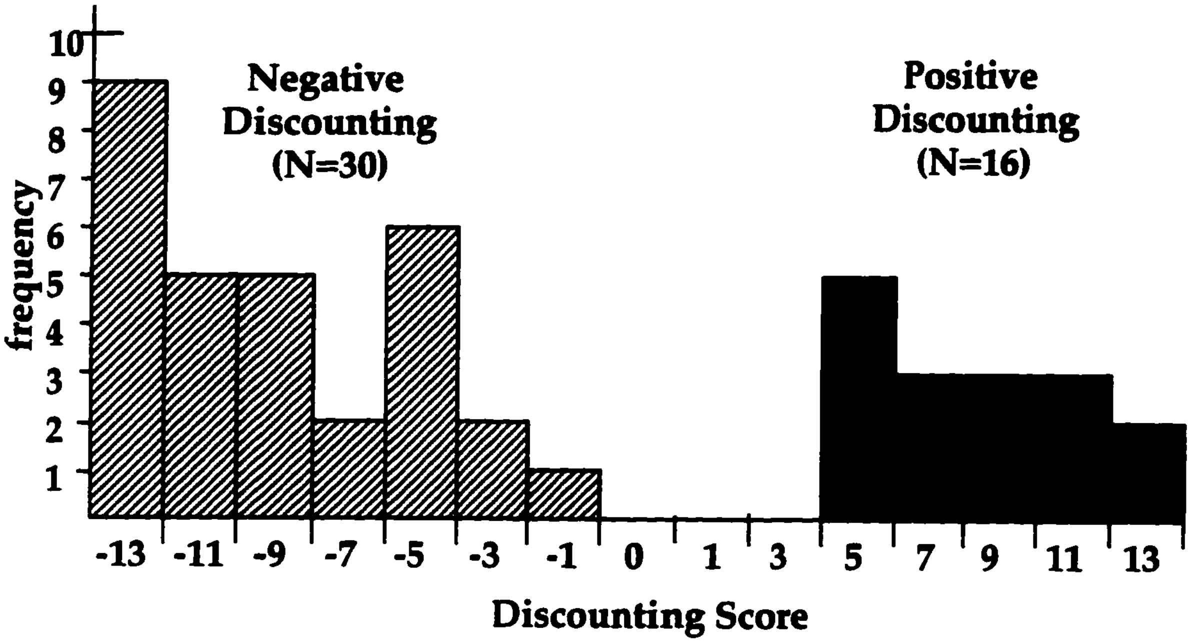


Figure 2: Frequency Distribution of the Discounting Score in Experiment 1

Each section of the horizontal axis represents two possible discounting scores. For example, the "-9" bar represents the number of subjects who had a discounting score of either -10 or -9. There are three exceptions to this: the section on the far left (-13), the section in the middle (0), and the section on the far right (13). These sections represent one score each.

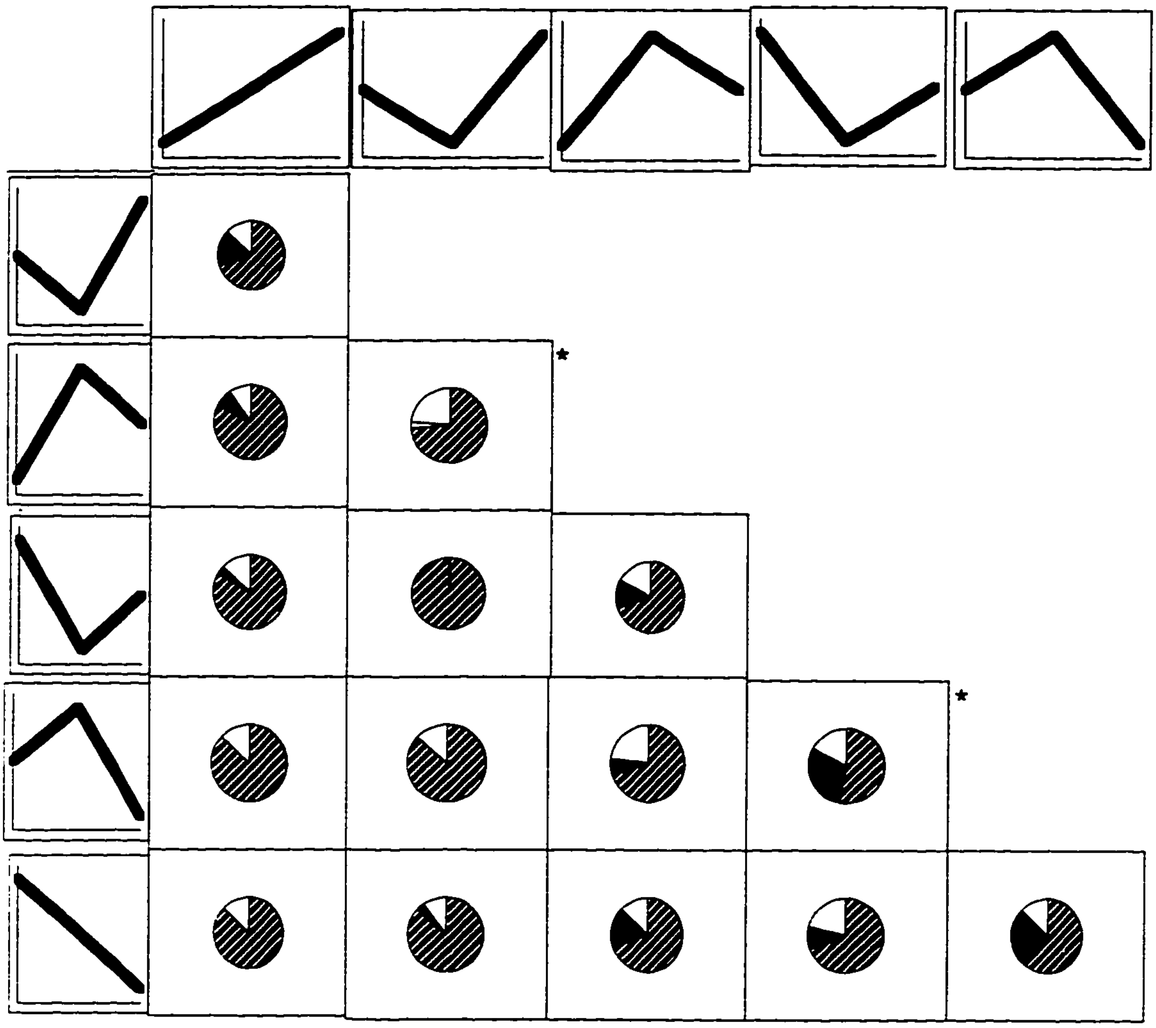


Figure 3: Preference Patterns of Negative Discounting Subjects in Experiment 1

proportion of NEGATIVE DISCOUNTING subjects who chose the column sequence over the row sequence twice

proportion of NEGATIVE DISCOUNTING subjects who chose the row sequence over the column sequence twice

proportion of NEGATIVE DISCOUNTING subjects who did not replicate choice

* For the pairs marked with an asterisk, neither positive nor negative discounting necessarily imply which sequence should be preferred.

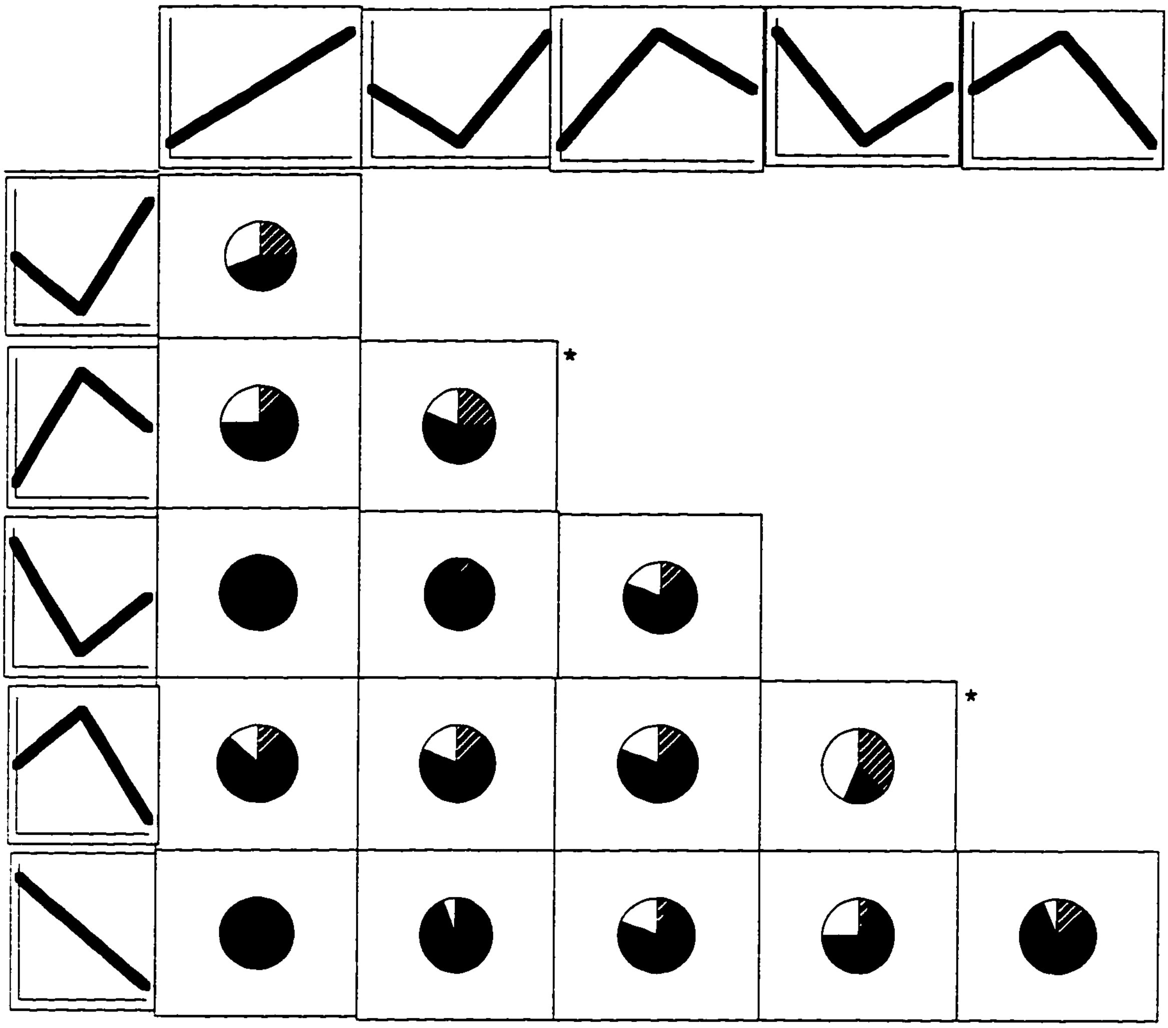


Figure 4: Preference Patterns of Positive Discounting Subjects in Experiment 1

▨ proportion of POSITIVE DISCOUNTING subjects who chose the column sequence over the row sequence twice

■ proportion of POSITIVE DISCOUNTING subjects who chose the row sequence over the column sequence twice

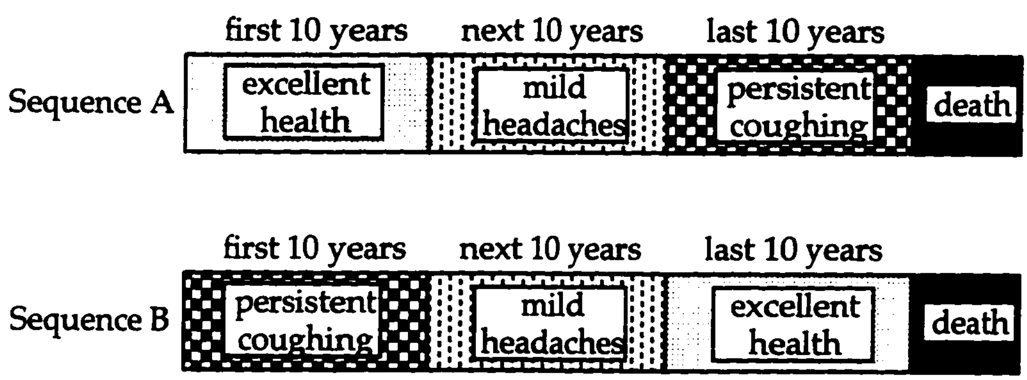
□ proportion of POSITIVE DISCOUNTING subjects who did not replicate choice

* For the pairs marked with an asterisk, neither positive nor negative discounting necessarily imply which sequence should be preferred.

CHAPTER 3: FOREKNOWLEDGE AND PREFERENCES BETWEEN FUTURE OUTCOMES

INTRODUCTION

In Experiment 1, the majority of subjects displayed negative discounting. In Experiment 2, we empirically tested one possible explanation for that finding: savoring and dread (Loewenstein and Prelec, 1993). To understand this explanation, consider the following two health sequences:



Notice that Sequence A is decreasing, whereas Sequence B is increasing. Positive discounting implies a preference for A, but negative discounting implies a preference for B. What would it be like to live through Sequence B? The first ten years contain the worst health state: persistent coughing. During those years, you may be comforted by the knowledge that you will experience better health in the upcoming 20 years. We refer to this as "savoring": current enjoyment of a future outcome.

What about Sequence A? Health is excellent during the first 10 years, but it worsens over time. One might feel dread about the future because it involves poorer health. Choosing B over A accomplishes two psychological goals simultaneously: it maximizes savoring and minimizes dread. Recall that B is the preferred sequence according to negative discounting. Therefore, savoring and dread can be used to explain negative time preference.

Loewenstein (1987) found that subjects preferred to postpone a kiss from their chosen movie star for three days rather than receive the kiss immediately, in three hours, or in one day. Those choices imply negative time preference. Also, however, subjects preferred to have the kiss in three days rather than in one year. For this particular outcome, pleasure from anticipation prevails up to a certain point, but after that point impatience prevails.

Loewenstein (1993) asked undergraduates three questions:

1.	Which would you prefer if both were free?	n=95
	A. Dinner at a fancy French restaurant	86%
	B. Dinner at a local Greek restaurant	14%
	For those who prefer French:	
2.	Which would you prefer?	n=82
	C. Dinner at the French restaurant on Friday in 1 month	80%
	D. Dinner at the French restaurant on Friday in 2 months	20%
3.	Which would you prefer?	n=82
	E. Dinner at the French restaurant on Friday in 1 month, and dinner at the Greek restaurant on Friday in 2 months	43%
	F. Dinner at the Greek restaurant on Friday in 1 month, and dinner at the French restaurant on Friday in 2 months	57%

The data for question 3 can be explained by savoring. 57% of the French-preferring subjects preferred to have the Greek dinner first and the French dinner afterwards. In other words, they chose the increasing sequence of dinners. Subjects may have wanted more time to savor the French dinner and so chose the increasing sequence.

The savoring/dread explanation requires an assumption which we call "foreknowledge". Specifically, in order to savor or dread future outcomes, the person must know beforehand what outcomes will occur. In Experiment 1, we did not specify to subjects whether or not they would have foreknowledge about future health states if they were to live through one of the sequences. They may have assumed foreknowledge, but there is no guarantee that they did. Instead, they may have assumed that each sequence would be lived in complete ignorance of future health states.

The design of Experiment 2 was guided by two observations. First, if the savoring/dread explanation of negative discounting is valid, then eliminating foreknowledge will reduce the extent of negative discounting. Without foreknowledge, neither savoring nor dread is possible. Second, if the savoring/dread explanation of negative discounting is valid, then emphasizing foreknowledge will increase the extent of negative discounting. With increased awareness of foreknowledge, it is reasonable to expect that the effects of savoring and dread will be enhanced.

Consequently, subjects were divided into groups that received different instructions regarding foreknowledge. Some subjects were told that they *should not* assume foreknowledge (the "No Foreknowledge" group, abbreviated as the "N" group). Other subjects were told that they *should* assume foreknowledge (the "Foreknowledge" group, abbreviated as the "F" group). Still other subjects received no instructions regarding foreknowledge (the "Unspecified" group, abbreviated as the "U" group). If the savoring/dread explanation is valid, then negative time preference should be least prevalent for the N group and most prevalent for the F group. The degree of negative time preference for the U group should be between these two extremes.

METHOD

SUBJECTS

86 University of Washington students participated. All were enrolled in Psychology courses and received extra credit for their participation.

HEALTH STATES

Three health states were used: excellent health, persistent coughing (abbreviated "pc"), and severe headaches (abbreviated "sh"). We used "severe" headaches instead of "mild" headaches because pilot data on the time tradeoff task indicated that many subjects were unwilling to give up any time with "mild headaches". The health state descriptions for persistent coughing and severe headaches are in Appendices A and K, respectively.

DESIGN AND MATERIALS

Each subject was randomly assigned to one of three groups: No foreknowledge, Unspecified, and Foreknowledge (hereafter abbreviated as "N", "U", and "F", respectively). The issue of foreknowledge was not discussed with the 29 subjects in the U group. The 29 subjects in the N group were explicitly told that all hypothetical events and occurrences in the experiment would be *unanticipated*: they should assume that they *would not* know in advance about future things that would happen to them. As an example, we described the idea of having a surprise conversation with a specific public person (see Appendix D). Because the conversation would be a surprise, the subject would not be able to plan in advance for it.

The 28 subjects in the F group were explicitly told that all hypothetical events and occurrences in the experiment would be *anticipated*: they should assume that they *would* know in advance about future things that would happen to them. We used the same example as for the N group, but we changed it so that the conversation would be fully anticipated and the subject could prepare in advance for it (see Appendix D).

Before each and every choice, subjects in the N and F groups were given an *explicit reminder* about the foreknowledge assumption that should be made when making that choice. This reminder was specially tailored to the choice situation. There were four choices situations in all: restaurant dinners, entertainment shows, time tradeoffs, and health sequences.

The restaurant questions were very similar to the questions asked by Loewenstein and Prelec (1993), and are listed in Appendix E. The entertainment questions (Appendix F) had the same structure as the restaurant questions, but the event of going out to dinner at a restaurant was replaced by going out to a popular show (either a theatrical play or a movie). In both of these question sets, the last question involved sequences: would you prefer an increasing or a decreasing sequence of these events?

Four time tradeoff indifference points were determined: two for severe headaches, and two for persistent coughing. For each health state, one indifference point was found using a 2-year duration, and another indifference point was found using a 10-year duration. A comparison of these two indifference points allows us to test the QALY model's assumption of constant proportional time tradeoff.

Regarding health sequences, subjects chose between a decreasing sequence and an increasing sequence. The sequences were presented in the same manner as in Experiment 1 (via shaded bar graphs). Each sequence was divided into 3 10-year intervals. The decreasing sequence began with excellent health, and then had persistent coughing, and then had severe headaches; the increasing sequence had the reverse order. Note that these sequences are "decreasing" and "increasing" only if the subject believed severe headaches to be worse than persistent coughing. Therefore, we created a second sequence choice by switching each "severe headaches" interval to a "persistent coughing" interval and by switching each "persistent coughing" interval to a "severe headaches" interval. This second choice was between an increasing sequence and a decreasing sequence for those subjects who stated that persistent coughing would be worse than severe headaches.

PROCEDURE

Separate general instructions were first given to subjects in the N and F groups regarding the foreknowledge assumption. Subjects then answered the four restaurant questions and the four entertainment questions. The order of these two question sets was determined randomly for each subject. Subjects in the N and F groups received specific instructions before answering these two sets of questions (see Appendix G). All subjects had been told to ignore scheduling conflicts when answering all questions.

After the restaurant questions and the entertainment questions, all subjects answered time tradeoff questions. The time tradeoff task was first explained in detail. Then, to provide an example of a time tradeoff judgment, the typical health limitations of having angina were described and subjects were asked whether they would prefer living for 60 years with angina or for 50 years without angina. After this example, subjects performed the four time tradeoff tasks. The foreknowledge instructions that were used during the time tradeoff task are listed in Appendix H. Each subject did the tasks in one of four random orders: (sh2, sh10, pc2, pc10), or (sh10, sh2, pc10, pc2), or (pc2, pc10, sh2, sh10), or (pc10, pc2, sh10, sh2).

In each task, the ping-pong method for finding an indifference point was used. In this method, the subject is first presented with the extreme ends of a choice, and then one option is altered back and forth until an indifference point is reached. For example, for the "sh10" time tradeoff task, the subject is first asked whether they would prefer living for 10 years in excellent health or living for 10 years with severe headaches. Of course, all subjects choose the former option. Then the subject compares "immediate death" to 10 years with severe headaches. Almost all subjects would choose the latter option. If so, then the subject compares 9.5 years in excellent health to 10 years with severe headaches. If the subject chooses the former, then an indifference point has been reached (here the indifference point would be assumed to be halfway between 9.5 and 10 years, or 9.75 years, which corresponds to a time-tradeoff utility of (9.75/10, or 0.975). But if the subject chooses the latter, the process continues with a comparison between 0.5 years in excellent health and 10 years with severe angina. This back-and-forth approach continues until an indifference point is reached. For the 2-year time period, the smallest time interval was 0.1 years instead of 0.5 years.

For all choices during the time tradeoff tasks, subjects were given a "c" option which they were told to choose only if options a and b were about the same in preference. If a subject chose this "c" option, then the current choice was taken as the inferred indifference point.

After the time tradeoff tasks, subjects were asked which health state would be worse: severe headaches or persistent coughing? Then they rated the size of this difference on a 1-5 scale where 1="slightly worse", 2="somewhat worse", 3="moderately worse", 4="much worse", and 5="very much worse".

The two health sequences choices were presented next. Subjects were told to imagine having an incurable illness in which they would have only 30 years of remaining life. The notion of a health sequence was described, and the foreknowledge instructions for groups N and F are in Appendix I.

The assumption question occurred at the end of the experiment, and it appears in Appendix J. It was included for two reasons: 1) to determine the proportion of subjects in the U group who assumed foreknowledge, and 2) to determine whether subjects in the N and F groups reported making the appropriate assumption throughout the experiment. All subjects in the N group should have chosen "b", whereas all subjects in the F group should have chosen "a". After answering this questions, subjects were debriefed and given credit for participating.

RESULTS

11 subjects (8 in the N group and 3 in the F group) inappropriately answered the assumption check question at the end of the

experiment. Despite repeated reminders of the instructions regarding the assumption of foreknowledge, these 11 subjects stated that they had assumed exactly the opposite of what they had been told to assume: the 8 subjects in the N group stated that they *had* assumed foreknowledge throughout, and the 3 subjects in the F group stated that they had assumed *no* foreknowledge throughout. Because these subjects may not have followed instructions, their data were removed. Therefore, there were 75 subjects in the analyses.

We had hypothesized that eliminating foreknowledge would make increasing sequences less attractive because future events could not be savored. The results for the dinner questions and the entertainment questions, however, did not support this hypothesis (see Table 1). In the upper half of that table, the proportion of subjects who chose the increasing sequence of dinners was actually slightly *higher* for the N group (7/21=0.33) than for the U group (7/29=0.24) and the F group (6/25=0.24), even though the N group was told to assume that they would not know about the dinners in advance. Also, notice that the majority of subjects preferred the decreasing sequence over the increasing sequence in all three groups, which suggests positive discounting.

The results for the entertainment questions (lower half of Table 1) are similar: the increasing sequence of shows was chosen equally often by subjects in the N group (8/21=0.38) and the U group (11/29=0.38) and slightly less often by subjects in the F group (5/25=0.20). In addition, most subjects preferred the decreasing sequence of shows over the increasing sequence of shows, again suggesting positive discounting. The resounding message of Table 1 is that the foreknowledge did not affect preferences as hypothesized.

Foreknowledge was also hypothesized to cause lower time tradeoff values: a person who knows the date of their death might try to "live life to the fullest" in their remaining years, and therefore might be willing to give up more life years in order to live in excellent health. The removal of the assumption of foreknowledge (the N Group), therefore, might result in higher time tradeoff values. But, as will be shown, the foreknowledge manipulation had little or no effect on time tradeoff values.

The time tradeoff data were analyzed in two separate ways. In the first analysis, all 75 subjects were included. In the second analysis, only 52 subjects were included. These 52 subjects were selected because all of their time tradeoff values were in the interval between 0 and 1 (exclusive). In a preliminary examination of the distribution of time tradeoff values, there was a surprisingly large number of subjects (23) who produced a time tradeoff value that was outside of this range. These responses may indeed be valid, but we suspect that a more likely possibility is that these subjects did not fully understand the time tradeoff task. For example, when presented with a choice between "immediate death" and "living for 10 years with persistent coughing", some subjects chose immediate death. This choice seems to be a rather severe reaction to how bad it would be to live with coughing, and it implies a negative time tradeoff value. Other subjects, when presented with a choice between "living for 10 years in excellent health" and "living for 10 years with persistent coughing", chose to live *with* the coughing rather than *without* the coughing. It is not clear why a person would choose in this way, and so we suspect that these subjects did not fully understand the task. However, this suggestion is based solely on our subjective judgment, and so the first analysis that includes all 75 subjects may be a more acceptable approach.

Figures 5, 6, 7, and 8 show means and confidence intervals of four time tradeoff values: severe headaches 2 years, severe headaches 10 years, persistent coughing 2 years, and persistent coughing 10 years. In each figure, the three points on the left represent the means with N=75, and the three points on the right represent the means with N=52. Note that, regardless of whether we used N=75 or N=52, the confidence intervals overlap in all four figures. The manipulation of foreknowledge assumptions, therefore, did not reliably affect time tradeoff values.

To analyze constant proportional time tradeoff, we combined the data from the three groups. There were two tests of constant proportional time tradeoff: one for severe headaches, and one for persistent coughing. The overall mean time tradeoffs for severe headaches were 0.586 for the 2-year time horizon and 0.555 for the 10-year time horizon. The 95% confidence interval around this difference of 0.031 had a lower bound of -0.025 and an upper bound of 0.087. Because this confidence interval includes 0, we conclude that constant proportional time tradeoff was not violated for severe headaches. For persistent coughing, the means were 0.672 for the 2-year time horizon and 0.662 for the 10-year time horizon. The difference of 0.010 had a 95% confidence interval lower bound of -0.042 and an upper bound of 0.061. Again, the effect was unreliable. When the results were recalculated for only the N=52 subset of subjects, still no reliable effects of constant proportional time tradeoff were found.

Next, we considered the possible influence of foreknowledge on sequence preferences in health. Perhaps removing foreknowledge lessens the attractiveness of the increasing sequence because future good health states cannot be anticipated and therefore cannot be savored. For each of the three groups, we computed the percentage of subjects who chose the increasing sequence of health over the decreasing sequence. These percentages were 52% (11/21) for the N group, 55% (16/29) for the U group, and 56% (14/25) for the F group. These percentages are in the hypothesized order, but the size of the differences are exceedingly small. It appears that the foreknowledge manipulation did not affect sequence preferences.

Interestingly, the data for health sequences suggested a slight tendency toward negative discounting, whereas the data for dinners and entertainment suggested strong positive discounting. Of the 41 subjects who preferred an increasing sequence of health, 27 of them chose the decreasing sequence of dinners (66%) and 23 of them chose the decreasing sequence of shows (56%). Therefore, even among subjects who used negative discounting for sequences, there was a tendency to use positive discounting for dinners and shows. Of the 34 subjects who preferred a decreasing sequence of health, a majority also preferred the decreasing sequence of dinners (28 of 34, or 82%) and shows (28 of 34, or 82%).

Finally, it is important to examine U subjects' responses to the assumption check question at the end of the experiment. Even though foreknowledge had not been mentioned to these 29 subjects, 20 of them (69%) stated that they had assumed foreknowledge whereas the other 9 (31%) stated that they had assumed no foreknowledge. This result suggests that foreknowledge was a common default assumption.

DISCUSSION

The savoring/dread explanation of negative discounting was not supported in Experiment 2. When foreknowledge was eliminated, most subjects (52%) chose the increasing health sequence over the decreasing health sequence. When foreknowledge was emphasized, a similar proportion of subjects (56%) chose the increasing health sequence. In other choice situations, the foreknowledge manipulation had little influence on preference patterns. Subjects tended to choose decreasing sequences of dinners and shows regardless of foreknowledge instructions. Time tradeoff values did not shift appreciably between groups.

It is possible that the foreknowledge manipulation was simply not strong enough to produce an effect. In other words, savoring/dread might still be a valid explanation of negative discounting, but Experiment 2 was an insufficient test of its predictions. We doubt this possibility. Clear, extensive instructions about foreknowledge were given to both the N and F groups (see again Appendices D, G, H, and I). Reminders about foreknowledge instructions appeared with *every single choice* in the experiment. The frequent reminders probably minimized memory loss about instructions. At the end of the experiment, subjects were asked what they had assumed about foreknowledge; those who answered inappropriately were removed from the dataset. There were, however, no other checks on subjects' understanding of the instructions. Even if they did understand the instructions, they may have failed to consider those instructions when making choices. All in all, we recognize the possibility that the manipulation was insufficient, but we doubt its plausibility.

Other than savoring/dread, Loewenstein and Prelec (1993) described two alternative explanations of negative discounting for sequences: adaptation/loss aversion, and recency. Neither of these explanations require foreknowledge, so they can be reconciled with the data in Experiment 2. In the adaptation/loss aversion explanation, the term "adaptation" addresses the idea that a person becomes accustomed to the current level of consumption and is sensitive to shifts from that level (Helson, 1964). Furthermore, people are particularly aversive to negative shifts, which are considered losses relative to the adaptation level (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991).

When evaluating a sequence, subjects imagine what it would be like to live through that sequence. For a decreasing sequence, subjects would adapt to the initial level of excellent health, and future health states would be viewed as losses relative to that initial level. In contrast, for an increasing sequence, the shifts are relative to initial poor health, and so they are viewed as gains. Because of loss aversion, the increasing sequence would be preferred. This explanation might involve foreknowledge, but it does not have to. A person could be completely unaware of future health states in the sequence, but they could still imagine themselves adapting to the current level and be averse to losses from that level.

In the recency explanation, the evaluation of a sequence is retrospective: when the sequence is over, how do you feel about it? From this perspective, the last interval in the sequence is likely to have the most influence (Ross and Simonson, 1991; Varey and Kahneman, 1992). Sequences that end well (e.g., increasing sequences) will generally be preferred to sequences that end poorly (e.g., decreasing sequences). There are cases, however, in which negative discounting does not necessarily imply that a happy-ending sequence will be preferred to an unhappy-ending sequence. Negative discounting gives more weight to later intervals, but early intervals still do receive some weight. Consider again the counterexample in Appendix C. One of the sequences ends in excellent health, and the other ends in moderate health. For some combinations of weights and utilities, negative discounting implies preference for the latter sequence even though its ending is worse.

One troublesome aspect of the data in Experiment 2 concerns the time tradeoff values. A large number of subjects (23 of 75) gave at least one extreme time tradeoff value. These subjects may not have fully understood the time tradeoff instructions. Unfortunately, the procedure during that task did not include any follow-up queries about peculiar responses. A better procedure would have included such queries. For example, suppose a subject stated that immediate death would be better than 10 years with persistent coughing. The task could have included a question immediately afterwards that asked the subject to verify the response. The subject would then have the option to change their answer. These questions may have reduced the number of extreme time tradeoff values, and therefore permitted a better test of group differences and also a better test of constant proportional time tradeoff.

Context does seem to influence the valence of discounting. For health sequences, a slight majority of subjects in all groups preferred the increasing health sequence, suggesting negative discounting. In contrast, for dinners and shows, a majority of subjects in all groups preferred the decreasing sequence, suggesting positive discounting. One reason for this finding is that the sequences were losses, whereas the dinners and shows were gains. As discussed previously, discounting is more likely to be negative for losses than for gains.

Table 1: Subject Counts for Dinner Questions and Entertainment Questions in Experiment 2

Dinner Questions	Preferred the decreasing sequence of dinners	Preferred the increasing sequence of dinners
N group (N=21)	14	7
U group (N=29)	22	7
F group (N=25)	19	6

Entertainment Questions	Preferred the decreasing sequence of shows	Preferred the increasing sequence of shows
N group (N=21)	13	8
U group (N=29)	18	11
F group (N=25)	20	5

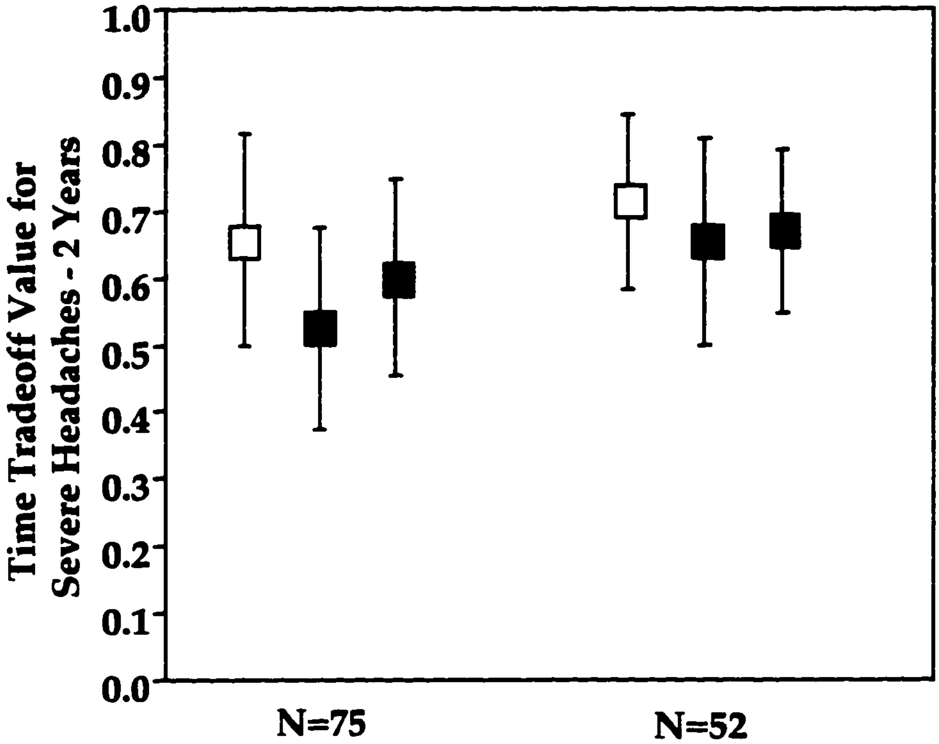


Figure 5: Time Tradeoff for Severe Headaches 2 Years in Experiment 2

□ N Group
■ U Group
■ F Group

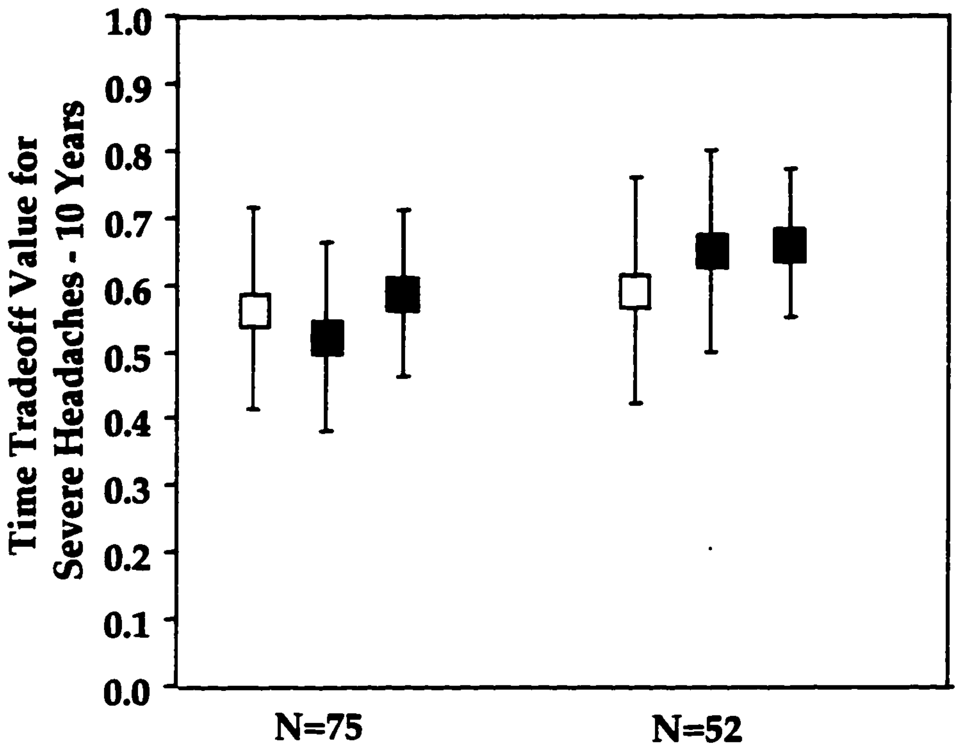


Figure 6: Time Tradeoff for Severe Headaches 10 Years in Experiment 2

□ N Group
■ U Group
■ F Group

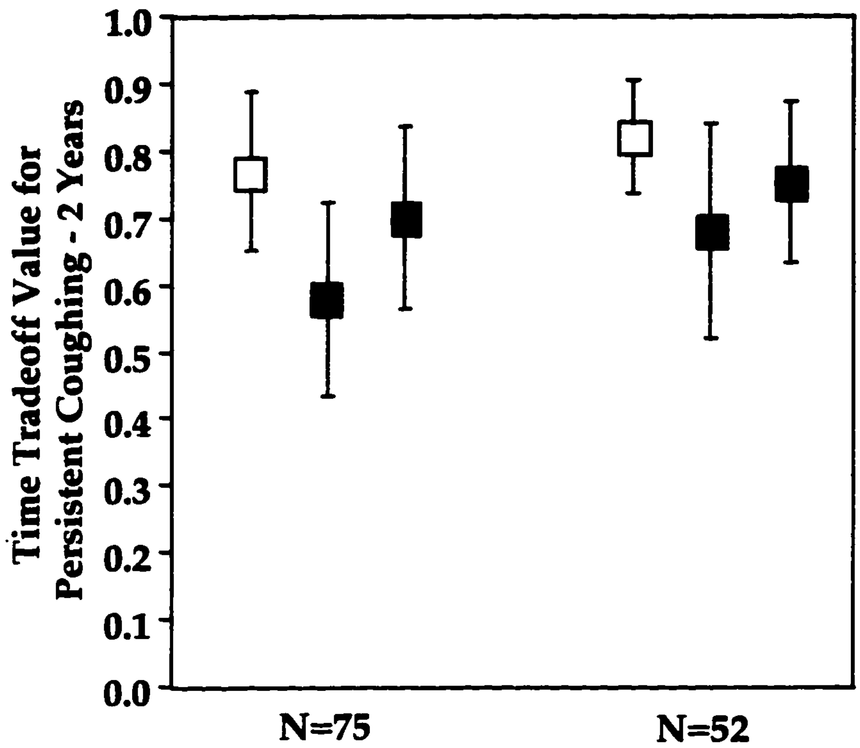


Figure 7: Time Tradeoff for Persistent Coughing 2 Years in Experiment 2

□ N Group
■ U Group
■ F Group

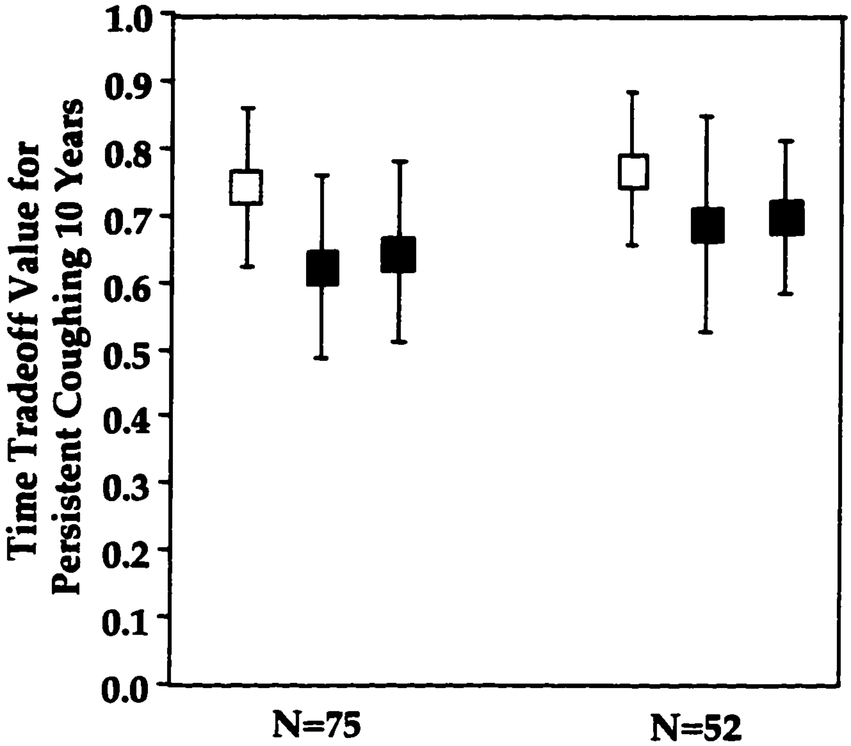


Figure 8: Time Tradeoff for Persistent Coughing 10 Years in Experiment 2

□ N Group
■ U Group
■ F Group

CHAPTER 4: INDEPENDENCE FOR EQUAL-INTERVAL HEALTH SEQUENCES

In Experiments 1 and 2, choices between health sequences revealed a tendency toward negative discounting. That inference, however, depends on the validity of the discounted QALY model for health sequences. That model was introduced in Equation 4:

$$\begin{aligned}&\textbf{Discounted } \mathit{QALYs}(\textbf{1st year in health state } Q_1\textbf{, then}\\ &\quad\textbf{2nd year in health state } Q_2\textbf{, then}\\ &\quad\ldots\\ &\quad\textbf{Yth year in health state } Q_Y) = \sum_{i=1}^{Y} \frac{H(Q_i)}{(1+d)^{i-1}}\end{aligned} \qquad \text{Equation 4}$$

According to this equation, the evaluation of a sequence can be modeled by the following process. Each interval in the sequence is first considered separately by computing the number of discounted QALYs associated with that interval. The overall value of the sequence is then computed as the sum of the separate intervals' discounted QALYs. Making a choice between two sequences involves determining which sequence has the higher overall value.

The discounted QALY model, however, might not be a good description of sequence evaluation. Notice that the structure of the model is additive; it is similar to the notion that the whole is equal to the sum of its parts. Subjects might instead evaluate sequences from a holistic perspective that cannot be incorporated into an additive model. If this were true, then a different model would be necessary.

What patterns of choice would suggest that subjects are taking a holistic perspective? One possibility concerns a preference for *spreading*: the degree to which the best (or worst) outcomes of a sequence are spread evenly over time. Subjects might systematically prefer a sequence whose best outcomes are temporally distant over a sequence whose best outcomes are bunched together. This preference might be based on the notion that the best outcomes have more impact and are more appreciated if they are separated in time. If those outcomes were consecutive, then perhaps they have less impact.

Some researchers have indeed found a preference for spreading (Linville and Fischer, 1991; Loewenstein and Prelec, 1993; Thaler and Johnson, 1990). Loewenstein and Prelec (1993) presented subjects with two choices:

Which would you prefer?

Option	This weekend	Next weekend	Two weekends from now	Choices
A	Fancy French	Eat at home	Fancy Lobster	54%
B	Eat at home	Fancy French	Fancy Lobster	46%
C	Fancy French	Eat at home	Eat at home	16%
D	Eat at home	Fancy French	Eat at home	84%

The typical subject chose to spread good outcomes (choosing A over B) and to spread bad outcomes (choosing D over C) ("eat at home" is bad relative to the other outcomes). A psychological justification for these choices might proceed as follows. Fancy dinners are special, and they are best enjoyed when separated in time, hence the choice of A over B. Eating at home is fairly routine, and so it would be better to break the routine by eating at a fancy French dinner during the middle weekend, hence the choice of D over C.

The above preference pattern cannot be explained by an additive model. To see why, suppose we used the following additive model:

$$\text{U(Dinner 1, then Dinner 2, then Dinner 3)} = \sum_{i=1}^{3} w_i U(\text{Dinner } i) \qquad \text{Equation 8}$$

where w_i is the weight associated with time interval i, and U(Dinner i) is the utility of dinner i. This model is structurally identical to the discounted QALY model in Equation 4.[10] In the above example, preferring A over B implies that

$$\begin{aligned} &w_1 U(\text{French}) + w_2 U(\text{home}) + w_3 U(\text{Lobster}) > \\ &w_1 U(\text{home}) + w_2 U(\text{French}) + w_3 U(\text{Lobster}) \end{aligned} \qquad \text{Equation 9}$$

Cancelling the common term from both sides,

$$w_1 U(\text{French}) + w_2 U(\text{home}) > w_1 U(\text{home}) + w_2 U(\text{French}) \qquad \text{Equation 10}$$

Preferring D over C implies that

$$\begin{aligned} &w_1 U(\text{French}) + w_2 U(\text{home}) + w_3 U(\text{home}) < \\ &w_1 U(\text{home}) + w_2 U(\text{French}) + w_3 U(\text{home}) \end{aligned} \qquad \text{Equation 11}$$

Cancelling the common term,

$$w_1 U(\text{French}) + w_2 U(\text{home}) < w_1 U(\text{home}) + w_2 U(\text{French}) \qquad \text{Equation 12}$$

Equations 10 and 12 are contradictory. It is not possible to choose a set of weights and utilities to account for this preference pattern. Loewenstein and Prelec (1993) devised a nonadditive model to address this problem.

---

[10] The isomorphism can be made clear by observing that $w_i = \dfrac{1}{(1+d)^{i-1}}$

The above pattern of preference is a violation of *independence*. Stated generally, independence is the requirement that if any two sequences have the same outcome during interval i, then preference between them does not switch if the outcome during interval i is changed (refer again to pages 6 and 7). The discounted QALY model for health sequences implies that independence should not be violated. A violation of independence for health sequences has been called a "sequence effect" (Krabbe and Bonsel, in press). This terminology highlights the idea that the valuation of a sequence can involve its global properties as well as its constituent parts.

Krabbe and Bonsel (in press) reported evidence for a sequence effect in health. They employed two variations on the time tradeoff task. In the first variation, subjects were presented with one health scenario in which they would live for 10 years in an intermediate health state (this state was defined precisely). They were asked to compare it to another scenario which would also involve 10 years of life, but those 10 years were divided into a "best imaginable health state" interval (occurring first) and "worst imaginable health state" interval (occurring afterwards). Note that this is a decreasing sequence. For the entire group of subjects, Krabbe and Bonsel (in press) determined what interval durations within the sequence (e.g., 6 years in best health, then 4 years in worst health) produced indifference between the sequence and 10 years in intermediate health.

The second variation of the time tradeoff task was identical except the sequence was increasing instead of decreasing (e.g., 4 years in worst health, then 6 years in best health). Therefore, they found two indifference points for the same intermediate health state: one based on a decreasing sequence, and another based on an increasing sequence. They adjusted each of these indifference points by applying a discount rate of 0.05. This adjustment was performed, supposedly, in order to account for time preference effects so that a sequence effect could be examined in isolation. They found that the two adjusted indifference points were reliably different, and then interpreted this finding as evidence for the existence of a sequence effect.

It is not clear, however, that the discount rate for health sequences is 0.05. Indeed, in our Experiments 1 and 2, we found evidence for negative discounting. If the subjects in Krabbe and Bonsel's (in press) experiment used any discount rate other than 0.05, then any conclusions about a hypothesized sequence effect would be influenced.

We chose to investigate the sequence effect in a different way: by directly testing the independence assumption of the discounted QALY model. If independence is violated, then a sequence effect exists. Conversely, if independence is satisfied, then a sequence effect does not exist. It is important to recognize that our tests of independence do not depend on the choice of a discount rate. A subject could be using positive discounting, or negative discounting, or zero discounting. Regardless, the results of these independence tests can be used to draw inferences about the hypothesized sequence effect.

METHOD

SUBJECTS

110 University of Washington students participated. All were enrolled in Psychology courses and received extra credit for their participation.

Health States

Three health states were used: excellent health, mild coughing, and severe headaches. Headaches were described as "severe", and coughing was described as "mild". These descriptions were made in order to increase agreement between subjects regarding which of the two intermediate health states would be worse. Given the health state descriptions (reproduced in Appendix K), it was expected that most subjects would feel that severe headaches would be worse than mild coughing.

Design and Materials

As in previous experiments, each health sequence lasted for 30 years, and the 30 years were divided into three ten-year intervals. Sequences were constructed in order to conduct 27 tests of independence (see Table 2). The presentation format of the sequences was identical to Experiments 1 and 2 (shaded bar graphs). Each test involved two choices: one in which the health state during the tied interval was excellent health, and another in which the health state during the tied interval was severe headaches. For example, the first row of Table 2 represents a test of independence in which the first choice is between "excellent health then excellent health then mild coughing" and "excellent health then mild coughing then excellent health". These two sequences are tied on the first interval. In the second choice, excellent health in the first interval has been changed to severe headaches for both sequences. Independence requires that preference is unaffected by this change.

Nine of the tests involved sequences that were tied on the first interval (tests #1-#9), nine involved sequences tied on the second interval (tests #10-#18), and nine involved sequences tied on the third interval (tests #19-27). Within each group of nine, the first three tests involved "permutations": in each pair of sequences, any sequence could be reordered to produce the other sequence of that pair. The next two tests involved "replacements": in each pair of sequences, one sequence contained excellent health and severe headaches in two of its intervals whereas the other sequence contained mild coughing in both of those intervals. The remaining four tests involved "unequal total health": one sequence in the pair had better total health than the other sequence. These four tests are least likely to produce violations of independence because a subject could choose solely on the basis of total health and thereby satisfy independence. Subjects' preferences for particular sequences of health (e.g., an increasing sequence), however, may be strong enough to overcome a difference in total health. Then, they might choose a favorable sequence of health states over an unfavorable sequence even if the latter has better total health.

Because of limitations on experiment time, each subject participated in 18 of the 27 tests. Each subject received one of six stimulus sets. The first set contained choices in which sequences were either tied on the first interval or tied on the second interval (tests #1-#18). In the second set, sequences were tied on either the first interval or the third interval (tests #1-#9 and #19-#27). In the third set, sequences were tied on either the second or third interval (tests #10-#27). Sets 4, 5, and 6 contained the same pairs as sets 1, 2, and 3, respectively, but the assignment of sequences to A and B slots was reversed (as in Experiment 1).

In addition to the choices involving independence, all subjects were given a choice between the following two sequences: "Excellent health then mild coughing then severe headaches" versus "Severe headaches then mild coughing then excellent health". For those subjects who felt that severe headaches was worse than mild coughing (the strong majority), this is a choice between a fully decreasing sequence and a fully increasing sequence.

This choice was included in an attempt to replicate the finding from previous experiments that subjects tend to prefer an increasing sequence of health over a decreasing sequence of health.

Therefore, 37 choices comprised the first half of each subject's data: two choices for each of 18 independence tests, plus an additional choice between a fully increasing and a fully decreasing sequence. These pairs were placed in random order. After the presentation of these 37 pairs, they were all presented a second time to investigate replicability. For these second presentations, pairs were re-randomized such that the first 19 pairs in the first set were reordered and shown first, and the second 18 pairs in the first set were reordered and shown next. This randomization scheme ensured that at least 18 choices intervened between the two presentations.

Procedure

As in Experiment 1, subjects were told to imagine having an incurable illness in which they would have only 30 years of remaining life. Subjects read brief descriptions of each health state, and were then asked which of the two intermediate health states (severe headaches or mild coughing) would be worse. It was emphasized in the instructions that only one of the three health states would be experienced in any ten-year interval. For each of 74 pairs of sequences (37 for the first set and 37 for the replicated set), subjects were asked which sequence would be preferred if they actually had the illness. After all 74 choices had been made, subjects were debriefed and given credit for participating.

RESULTS

98 of the 110 subjects stated that severe headaches would be worse than mild coughing. Because the stimuli were created with the

assumption that severe headaches would be worse, only these 98 subjects were included in data analyses. Of these 98 subjects, 33 chose between sequences that were tied on either the first or second interval, 32 chose between sequences that were tied on either the first or third interval, and 33 chose between sequences that were tied on either the second or third interval. The independence tests pertaining to sequences tied on the first interval were based on data from 65 subjects (33 from the first group and 32 from the second group)[11]. Similarly, the independence tests pertaining to sequences tied on the second interval were based on 66 subjects' data[12], and those tests pertaining to sequences tied on the third interval were based on 65 subjects' data.

Figures 9, 10, and 11 display the results for the tests pertaining to sequences tied on the first, second, and third intervals, respectively. For all 27 tests of independence, the proportion of subjects who satisfied independence twice (dark grey area of each pie) was higher than the proportion of subjects who violated independence twice (light grey area of each pie). The tables to the right of the pies indicate the choice patterns of subjects in the dark grey and light grey areas of the pies.

The largest number of violations occurred for test #14 in which 13 subjects (20%) chose the sequence "SEE" over "MEM" but also chose "MSM" over "SSE". For this same test, however, there were 30 subjects (46%) who satisfied independence twice.

---

[11] One subject in the second group skipped one question (pertaining to independence test #5), and so the number of subjects for that test was 64 instead of 65.

[12] One subject in the first group skipped one question (pertaining to independence test #14), and so the number of subjects for that test was 65 instead of 66.

The white area in some of the pies might seem surprisingly large. Recall, however, that each independence test for a particular subject represents four choices: two presentations for each of two pairs of sequences. All subjects who switched preference between presentations for *either* the first pair *or* the second pair are represented by the white area of the pie.

We computed an "independence score" for each test for each subject. Each test was coded as 1 if independence was satisfied twice by that subject, 0 if independence was violated twice by that subject, and 0.5 if the subject did not replicate one or both of their choices. The mean independence scores (and 95% confidence intervals) for the 27 tests are shown in Figure 12. All of the 27 means were reliably higher than 0.5, indicating satisfaction of independence for all 27 independence tests.

To determine whether there was a sizable minority of subjects who consistently violated independence, we computed the sum of each subject's independence scores over 18 tests. The sum ranges from 0 (indicating maximum violation of independence) to 18 (indicating maximum satisfaction of independence). A score of 9 indicates neither consistent violation nor consistent satisfaction of independence.

A frequency distribution of the sums is shown in Figure 13. From this figure, it is clear that there was *not* a subgroup of subjects who tended to violate independence. Instead, 97 of the 98 subjects had a sum that was greater than 9. The mean was 14.1, and the 95% confidence interval around this mean had a lower bound of 13.6 and an upper bound of 14.6. There was one subject who consistently violated independence, represented by the light grey box on the left in Figure 13.

All subjects were asked to choose between a fully increasing sequence and a fully decreasing sequence. The results for this choice are very similar to those from Experiment 1: 57 subjects chose the increasing sequence twice (58%), 36 subjects chose the decreasing sequence twice (37%), and 5 subjects did not replicate their choice (5%).

The data from the permutations (tests #1-#3, #10-#12, and #19-#21) can be used to determine whether subjects split into a negative discounting group and a positive discounting group as they did in Experiment 1.[13] Each subject made 24 choices between permutation sequences, and so the discounting score ranged from -12 (indicating maximum negative discounting) to 12 (indicating maximum positive discounting). The frequency distribution of this score is in Figure 14. The distribution not as obviously bimodal as the distribution in Figure 2. Perhaps "positively skewed" would be a better description: many subjects had strong negative scores, and a smaller number of subjects had weaker negative scores and positive scores.

DISCUSSION

The primary result of Experiment 3 was unambiguous: subjects satisfied independence. Regardless of whether two sequences were tied on the first, second, or third interval, the choice between them did not depend on the level of health during the tied interval. We interpret this as evidence against a sequence effect. Choices between health sequences can be accurately modeled by discounted QALYs. The inferences about the valence of discounting in Experiments 1 and 2 still hold.

A standard tactic in the literature on judgment and decision making is to present a simple, widely-accepted model, and then criticize it by

---

[13] The sequence pairs involving replacements and unequal total health cannot be used to infer the valence of discounting.

reporting descriptive violations of its assumptions (Allais, 1953; Ellsberg, 1961; Kahneman and Tversky, 1979). Then one proposes a more complicated model that actually *predicts* the findings previously thought to be anomalous. In Experiment 3, however, we found relatively few violations of the independence assumption of the discounted QALY model. A more complicated model is not needed. This is good news from the standpoint of practicality: efforts to model sequence preferences do not need to incorporate a special sequence effect. Instead, the valuation of a sequence can be estimated by adding the valuations of its intervals.

Experiment 3 did not replicate the finding from Experiment 1 that subjects split into a negative discounting group and a positive discounting group. There were subjects who did clearly fall into each of these categories, but there were also subjects who fell somewhere in the middle (refer again to Figure 14). Experiment 3 did, however, replicate the finding from Experiments 1 and 2 that negative discounting is more common than positive discounting. Therefore, we have additional evidence that people tend toward negative discounting when choosing between health sequences.

Table 2: Independence Tests in Experiment 3

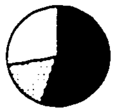
"E" denotes "excellent health; "S" denotes "severe headaches"; "M" denotes "mild coughing"; "iff" denotes "if and only if".

1)	E	E	M	>	E	M	E	iff	S	E	M	>	S	M	E
2)	E	E	S	>	E	S	E	iff	S	E	S	>	S	S	E
3)	E	M	S	>	E	S	M	iff	S	M	S	>	S	S	M
4)	E	E	S	>	E	M	M	iff	S	E	S	>	S	M	M
5)	E	M	M	>	E	S	E	iff	S	M	M	>	S	S	E
6)	E	E	M	>	E	S	E	iff	S	E	M	>	S	S	E
7)	E	E	S	>	E	M	E	iff	S	E	S	>	S	M	E
8)	E	E	S	>	E	S	M	iff	S	E	S	>	S	S	M
9)	E	M	S	>	E	S	E	iff	S	M	S	>	S	S	E
10)	E	E	M	>	M	E	E	iff	E	S	M	>	M	S	E
11)	E	E	S	>	S	E	E	iff	E	S	S	>	S	S	E
12)	M	E	S	>	S	E	M	iff	M	S	S	>	S	S	M
13)	E	E	S	>	M	E	M	iff	E	S	S	>	M	S	M
14)	M	E	M	>	S	E	E	iff	M	S	M	>	S	S	E
15)	E	E	M	>	S	E	E	iff	E	S	M	>	S	S	E
16)	E	E	S	>	M	E	E	iff	E	S	S	>	M	S	E
17)	E	E	S	>	S	E	M	iff	E	S	S	>	S	S	M
18)	M	E	S	>	S	E	E	iff	M	S	S	>	S	S	E
19)	E	M	E	>	M	E	E	iff	E	M	S	>	M	E	S
20)	E	S	E	>	S	E	E	iff	E	S	S	>	S	E	S
21)	M	S	E	>	S	M	E	iff	M	S	S	>	S	M	S
22)	E	S	E	>	M	M	E	iff	E	S	S	>	M	M	S
23)	M	M	E	>	S	E	E	iff	M	M	S	>	S	E	S
24)	E	M	E	>	S	E	E	iff	E	M	S	>	S	E	S
25)	E	S	E	>	M	E	E	iff	E	S	S	>	M	E	S
26)	E	S	E	>	S	M	E	iff	E	S	S	>	S	M	S
27)	M	S	E	>	S	E	E	iff	M	S	S	>	S	E	S

test #1 — A>B iff A'>B' — EEM>EME iff SEM>SME

	choose A'	choose B'
choose A	12 ■	7 ▧
choose B	1 ▧	27 ■

test #2 — A>B iff A'>B' — EES>ESE iff SES>SSE



	choose A'	choose B'
choose A	12 ■	5 ▧
choose B	6 ▧	24 ■

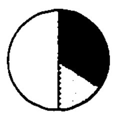
test #3 — A>B iff A'>B' — EMS>ESM iff SMS>SSM

	choose A'	choose B'
choose A	11 ■	3 ▧
choose B	4 ▧	28 ■

test #4 — A>B iff A'>B' — EES>EMM iff SES>SMM

	choose A'	choose B'
choose A	6 ■	10 ▧
choose B	1 ▧	31 ■

test #5 — A>B iff A'>B' — EMM>ESE iff SMM>SSE



	choose A'	choose B'
choose A	16 ■	0 ▧
choose B	10 ▧	6 ■

test #6 — A>B iff A'>B' — EEM>ESE iff SEM>SSE

	choose A'	choose B'
choose A	56 ■	2 ▧
choose B	0 ▧	0 ■

test #7 — A>B iff A'>B' — EES>EME iff SES>SME

	choose A'	choose B'
choose A	1 ■	3 ▧
choose B	0 ▧	56 ■

test #8 — A>B iff A'>B' — EES>ESM iff SES>SSM

	choose A'	choose B'
choose A	33 ■	2 ▧
choose B	3 ▧	5 ■

test #9 — A>B iff A'>B' — EMS>ESE iff SMS>SSE

	choose A'	choose B'
choose A	0 ■	1 ▧
choose B	2 ▧	46 ■

Figure 9: Independence for Sequences Tied on First Interval in Experiment 3

■ **subjects who satisfied independence twice**

□ **subjects who violated independence twice**

□ **subjects who did not replicate choices**

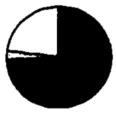
test #10 — A>B iff A'>B' — EEM>MEE iff ESM>MSE

	choose A'	choose B'
choose A	20	0
choose B	0	33

test #11 — A>B iff A'>B' — EES>SEE iff ESS>SSE

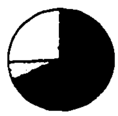
	choose A'	choose B'
choose A	16	0
choose B	3	40

test #12 — A>B iff A'>B' — MES>SEM iff MSS>SSM



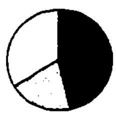
	choose A'	choose B'
choose A	14	0
choose B	2	35

test #13 — A>B iff A'>B' — EES>MEM iff ESS>MSM



	choose A'	choose B'
choose A	13	4
choose B	0	32

test #14 — A>B iff A'>B' — MEM>SEE iff MSM>SSE

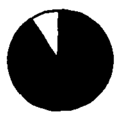


	choose A'	choose B'
choose A	15	0
choose B	13	15

test #15 — A>B iff A'>B' — EEM>SEE iff ESM>SSE

	choose A'	choose B'
choose A	40	0
choose B	2	6

test #16 — A>B iff A'>B' — EES>MEE iff ESS>MSE



	choose A'	choose B'
choose A	2	0
choose B	0	58

test #17 — A>B iff A'>B' — EES>SEM iff ESS>SSM

	choose A'	choose B'
choose A	34	0
choose B	1	8

test #18 — A>B iff A'>B' — MES>SEE iff MSS>SSE

	choose A'	choose B'
choose A	3	0
choose B	0	53

Figure 10: Independence for Sequences Tied on Second Interval in Experiment 3

■ **subjects who satisfied independence twice**

▒ **subjects who violated independence twice**

□ **subjects who did not replicate choices**

test #19 — A>B iff A'>B' — EME>MEE iff EMS>MES

	choose A'	choose B'
choose A	11	1
choose B	11	17

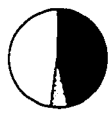
test #20 — A>B iff A'>B' — ESE>SEE iff ESS>SES

	choose A'	choose B'
choose A	9	3
choose B	7	22

test #21 — A>B iff A'>B' — MSE>SME iff MSS>SMS

	choose A'	choose B'
choose A	6	3
choose B	6	30

test #22 — A>B iff A'>B' — ESE>MME iff ESS>MMS



	choose A'	choose B'
choose A	8	2
choose B	2	22

test #23 — A>B iff A'>B' — MME>SEE iff MMS>SES

	choose A'	choose B'
choose A	18	0
choose B	9	15

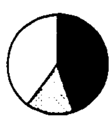
test #24 — A>B iff A'>B' — EME>SEE iff EMS>SES

	choose A'	choose B'
choose A	42	0
choose B	7	1

test #25 — A>B iff A'>B' — ESE>MEE iff ESS>MES

	choose A'	choose B'
choose A	1	0
choose B	2	56

test #26 — A>B iff A'>B' — ESE>SME iff ESS>SMS



	choose A'	choose B'
choose A	29	8
choose B	2	0

test #27 — A>B iff A'>B' — MSE>SEE iff MSS>SES

	choose A'	choose B'
choose A	0	0
choose B	3	53

Figure 11: Independence for Sequences Tied on Third Interval in Experiment 3

subjects who satisfied independence twice

subjects who violated independence twice

subjects who did not replicate choices

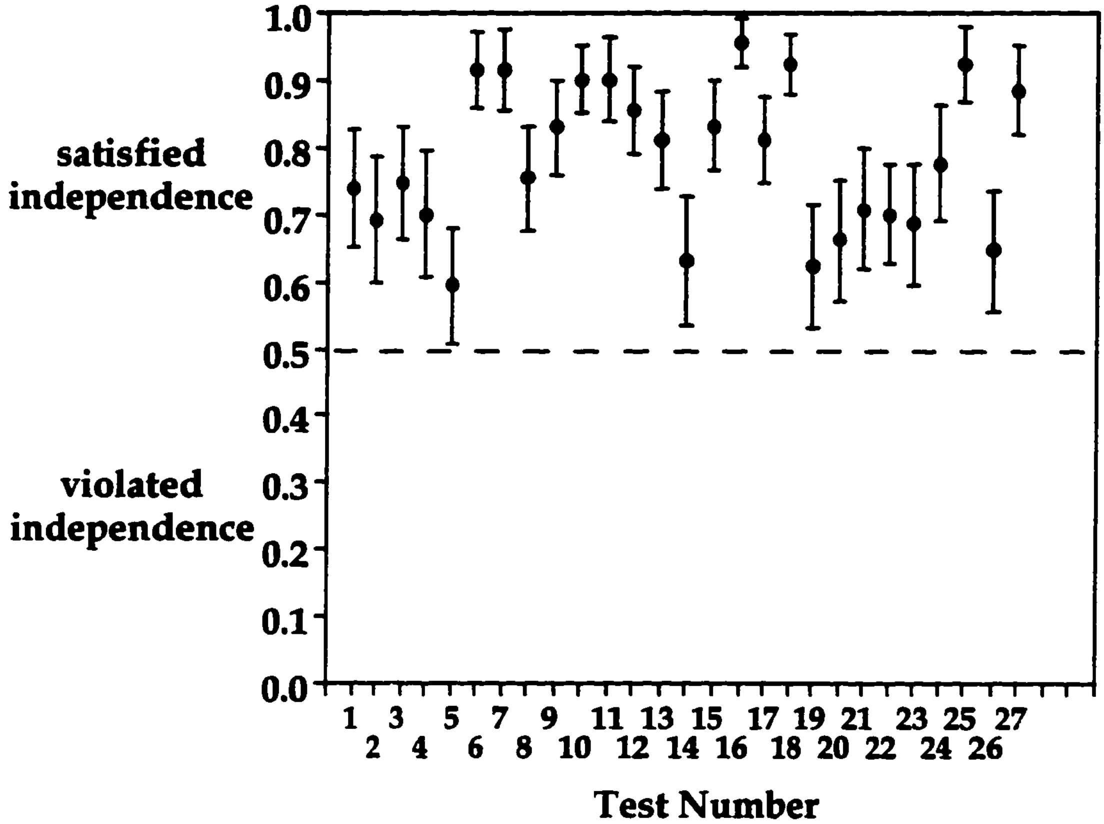


Figure 12: Means and 95% Confidence Intervals for Independence Tests in Experiment 3

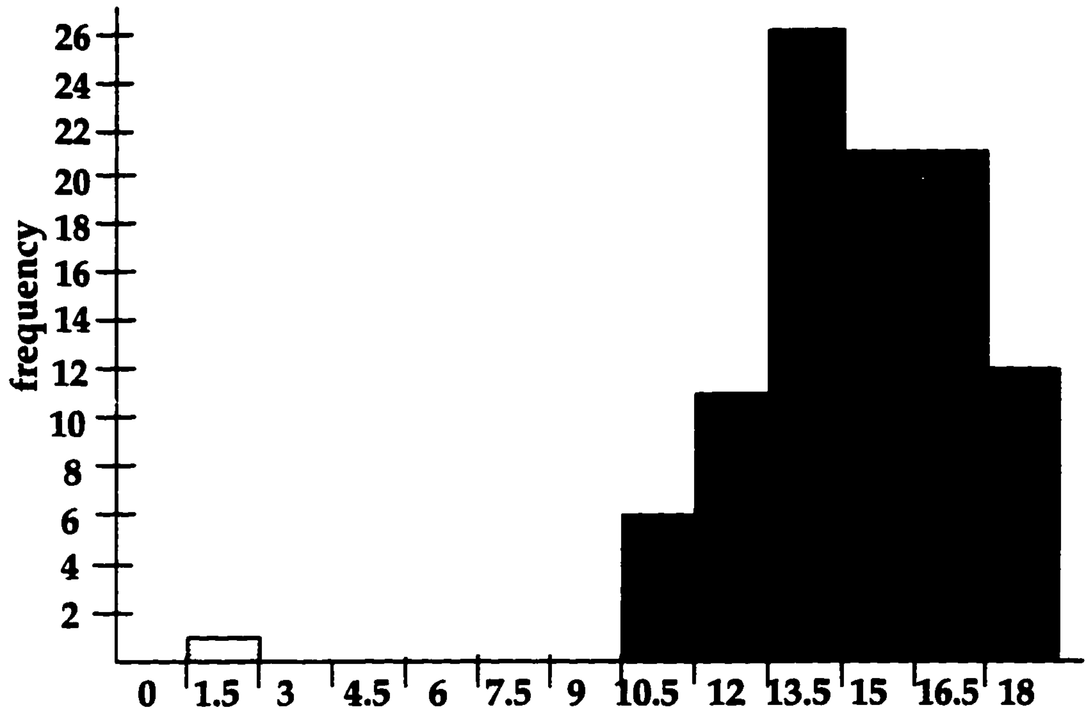


Figure 13: Frequency Distribution of the Independence Score in Experiment 3

Each section of the horizontal axis represents three possible independence scores. For example, the "11" column represents scores of 11, 11.5, and 12. There is one exception to this: the section in the middle (9) represents only a score of 9.

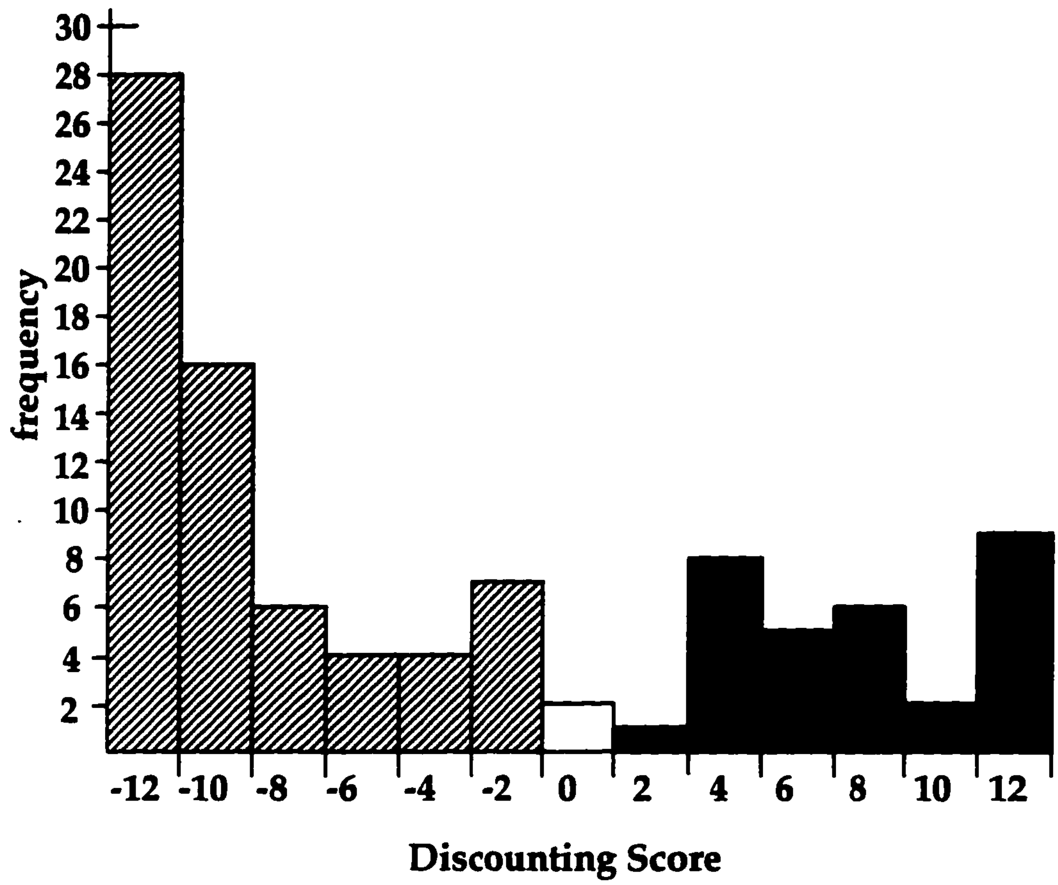


Figure 14: Frequency Distribution of the Discounting Score in Experiment 3

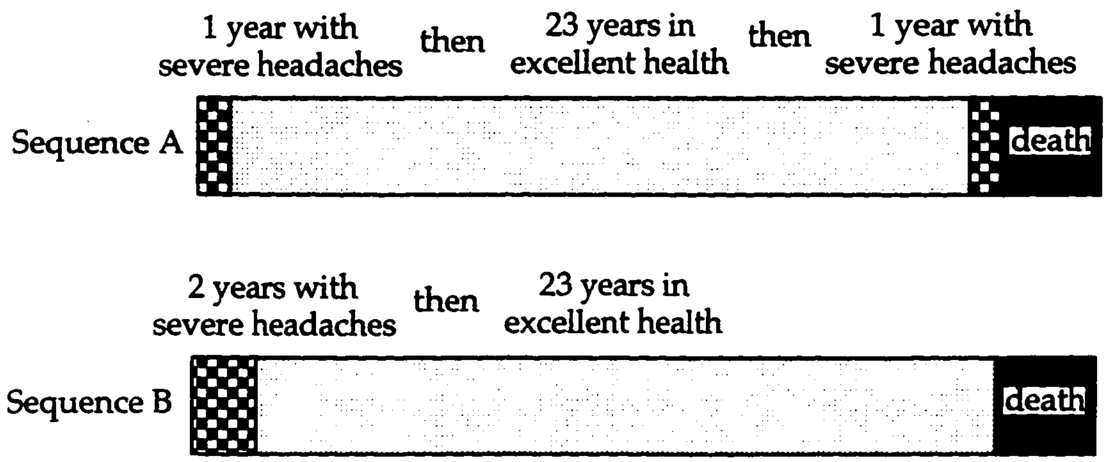
Each section of the horizontal axis represents two possible discounting scores, as in Figure 2. There is one exception to this: the section in the middle (0) represents only a score of 0.

CHAPTER 5: INDEPENDENCE FOR UNEQUAL-INTERVAL HEALTH SEQUENCES

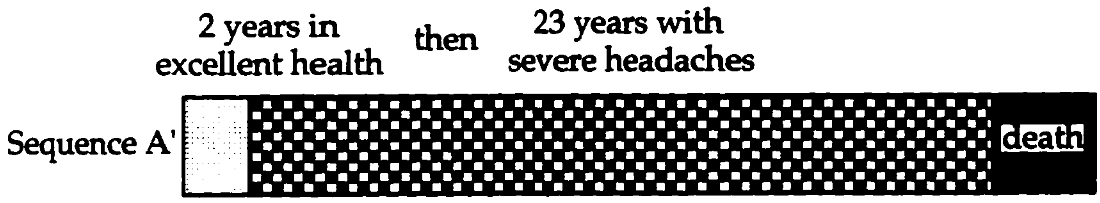
In Experiment 3, all of the independence tests involved sequences whose intervals had equal durations (10 years). This constant may have contributed to the widespread satisfaction of independence. In Experiment 4, therefore, subjects made choices between sequences of unequal-duration intervals. The independence tests were specially designed to be more sensitive to violations.

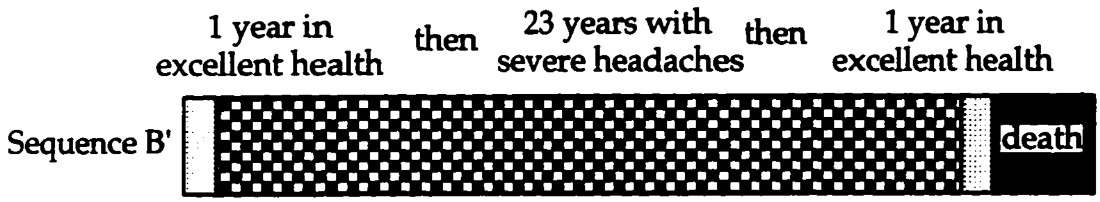
Specifically, we created sequences to accentuate the degree of spread between identical intervals. If subjects do have a preference for spreading, then that preference is more likely to be observed if spreading is maximized. A systematic preference for spreading violates independence (as shown by the three-dinner sequences introduced by Loewenstein and Prelec (1993) and described in our introduction to Experiment 3).

For example, one pair of sequences in Experiment 4 was:



Notice that these two sequences have the same health state during year 1, and also during years 3-24. We changed the health states during these tied intervals to create a corresponding pair for testing independence:





Independence requires that A is preferred to B if and only if A′ if preferred to B′. Preference for spreading implies that A will be chosen over B because A spreads the two years with severe headaches. Also, it implies that B′ will be chosen over A′ because B′ spreads the two years in excellent health. That pattern of preference violates independence.

Another novel aspect of Experiment 4 is the use of a very short-term health sequence in addition to the long-term health sequences (e.g., 30 years) already considered. Each short-term sequence lasted for one week, and at the end of the week the subject would return to normal health. On each day of the week, there would be either a painful procedure (to treat a hypothetical serious illness), or there would be no painful procedure. We hypothesized that this short-term scenario would be more likely to reveal a preference for spreading: subjects can best endure painful procedures if they are separated in time. As with the long-term sequences, a systematic preference for spreading will result in an independence violation.

METHOD

Subjects

65 University of Washington students participated. All were enrolled in Psychology courses and received extra credit for their participation.

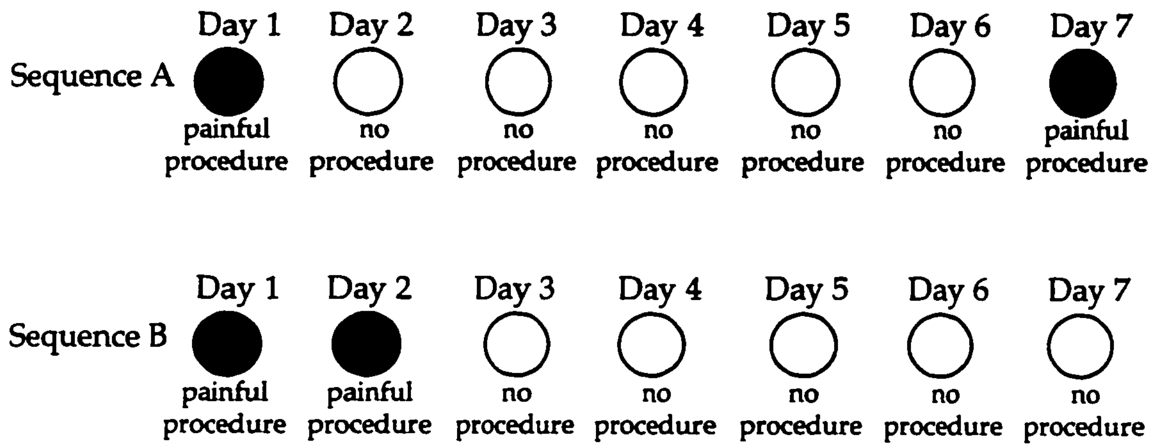
Health States

We used the same three health states (and associated descriptions) as in Experiment 3: excellent health, mild coughing, and severe headaches. In addition, a different health scenario was used in which subjects imagined having a serious but treatable illness. The required treatment, however, involved a painful medical procedure. The description of this scenario appears in Appendix L.

Design and Materials

There were three sets of choices in this experiment: the "headache" set, the "coughing" set, and the "painful procedure" set. For the first two sets, all sequences lasted for 25 years (ending in death), whereas for the painful procedure set, all sequences lasted for seven days (followed by a return to normal health).

In the headache set, all choices were between sequences whose only health states were excellent health and severe headaches (e.g., 23 years of excellent health followed by 2 years of severe headaches, versus 2 years of severe headaches followed by 23 years of excellent health). Similarly, in the coughing set, all sequences involved only excellent health and mild coughing. The presentation format of both the severe headaches set and the mild coughing set was identical to the examples given in the Introduction. Lastly, in the painful procedure set, all sequences involved a seven-day interval in which there was a painful procedure on some days but no procedure on other days. The presentation format of the painful procedure set was as below:



For each of the three sets, five tests of independence were performed (see Table 3). All sequences consisted of seven time intervals. For the headache set and the coughing set, five of the intervals lasted for one year (intervals 1,2,4,6, and 7) and two intervals lasted for ten years (intervals 3 and 5). For example, in the first choice of the first independence test in Table 3, the first alternative is a sequence in which the first six time intervals are lived in excellent health, and the last interval is lived with severe headaches. That sequence involves living for the first 24 years in excellent health and then living for the last year with severe headaches. The subject must compare that sequence to one in which the year of severe headaches occurs in the middle of the 25 year period instead of at the end.

Notice that those two sequences have the same health state during five of the seven intervals (1, 2, 3, 5, and 6). If independence is satisfied, then preference between those two sequences should not depend on the health state during the tied intervals. Therefore, if the health state during those intervals is changed from "excellent health" to "severe headaches" (see the second choice of the first independence test in Table 3), preference should not switch.

When a 25-year sequence was presented for the subject's consideration, the description combined adjacent intervals that contained the same health state. For example, the sequence "E E E E E E SH" was described as "24 years in excellent health, then 1 year with severe headaches" and the sequence "E E E SH E E E" was described as "12 years in excellent health, then 1 year with severe headaches, then 12 years in excellent health".

For the painful procedure set of choices, each of the seven time intervals lasted for one day. Notice, however, that the five independence tests in the painful procedure set were structurally identical to the five tests in the headaches set and also to the five tests in the coughing set. Each interval in "excellent health" was replaced by a day with "no procedure", and each interval in an intermediate health state was replaced by a day with a "painful procedure". Therefore, the fifteen independence tests in Experiment 4 can be viewed as five tests in each of three different contexts.

Independence tests 5, 10, and 15 involved choices between an increasing sequence and a decreasing sequence. In addition to testing independence, these choices were included in an attempt to replicate previous findings that subjects tend to prefer an increasing sequence.

In contrast to subjects in Experiment 3, each subject in Experiment 4 participated in all independence tests. For each independence test, both of the choices were presented twice to investigate replicability, and the assignment of sequences to either the A slot or the B slot was determined randomly and separately for each subject. In all, each subject made 60 choices (3 sets x 5 tests x 4 choices).

For each subject, the first 20 choices were all in the same set (e.g., the headache set), and the next 20 choices were in another set (e.g., the coughing set), and the last 20 choices were in the other set (e.g., the painful procedure set). The order of presentation of the three sets was determined randomly and separately for each subject (3!=six possible orders). Within each set of 20 choices, the first 10 were randomly ordered and presented. Then, the first five of these were randomized anew and presented as the first five replications. Finally, the other five choices were randomized anew and presented as the second five replications. This process ensured that at least five choices intervened between the two presentations.

PROCEDURE

First, the subject was randomly assigned to one of the six orders of presentation of the three sets. An introductory description of a health scenario was then given for the first set. For the headache and coughing sets, subjects were told to imagine having an incurable illness in which they would have only 25 years of remaining life. For the painful procedure set, subjects were told to imagine having a serious illness that must be treated in the upcoming week. Then the relevant health state for that set was described. After the health state description, an example of a choice was presented: 1 year in state X then 24 years in state Y, versus 24 years in state Y then 1 year in state X. For the headache and coughing sets, state X was the relevant health state for that set, whereas State Y was "excellent health". For the painful procedure set, the time scale was seven days instead of 25 years, and states X and Y were "painful procedure" and "no procedure", respectively.

The 20 choices in the first set were then presented; for each pair of sequences, the subject was asked which sequence would be preferred if they actually had the illness. After the completion of the 20 choices, the introductory description for the next set appeared. The subject then progressed through the health state description, and 20 choices in that set. Then the third and final set (along with relevant descriptions) was shown.

After the 20 choices of the third set, the subject was asked three questions about the general approach they had used during the experiment. These questions appear in Appendix M. There was one question for each of the three sets, and subjects answered these questions in the same order in which the sets had been presented earlier. For each question, strategy X emphasized the advantage in having poor health sooner rather than later, whereas strategy Y emphasized the advantage in having poor health later rather than sooner. The strategies were worded in an attempt to make them seem equally reasonable. After these questions had been answered, subjects were debriefed and given credit for their participation.

RESULTS

As in Experiment 3, a pie chart will be used for each test of independence to indicate three proportions: 1) the proportion of subjects who satisfied independence twice (dark grey area of each pie), 2) the proportion of subjects who violated independence twice (light grey area of each pie), and 3) the proportion of subjects who did not replicate one or both of their choices (white area of each pie). Also, the tables to the right of the pies indicate the choice patterns of subjects in the dark grey and light grey areas of the pies. The pie charts and tables are displayed in Figures 15 (severe headaches), 16 (mild coughing), and 17 (painful procedure).

There seems to be generally more violations of independence than in Experiment 3. For three tests in particular (#6, #7, and #11), the dark grey and light grey areas are approximately equal in size. These three tests will now be considered in detail.

For test #6, subjects did not show a consistent pattern of violations. Nine subjects chose EEEEEEM over EEEMEEE and also chose MMMMMME over MMMEMMM. These subjects, therefore, violated independence in such a way that they tended to choose sequences in which there was a long interval of constant health. There were 10 subjects, however, who violated independence in the opposite manner (EEEMEEE chosen over EEEEEEM, and MMMEMMM chosen over MMMMMME). That pattern indicates a preference for spreading.

In contrast, there was a consistent pattern of violation for test #7. Seventeen subjects (26%) chose MEEEEEE over EEEMEEE and also chose EMMMMMM over MMMEMMM, but there were only 5 subjects (8%) who violated independence in the opposite manner. For this test, most violators chose sequences in which there was a long interval of constant health.

In test #11, however, the more consistent pattern of violation was to avoid those kinds of sequences. 22 subjects (34%) chose NNNPNNN over NNNNNNP and also chose PPPNPPP over PPPPPPN, whereas only 6 showed the opposite pattern of violation. Most violators, therefore, preferred to have a day of novelty in the middle of a seven-day sequence rather than at the end. This might be viewed as a day of rest within a long series of painful procedure days (PPPNPPP), or as a painful day that is buffered by pain-free days preceding and following (NNNPNNN).

Tests #5, #10, and #15 contain choices between a fully increasing sequence and a fully decreasing sequence. As in previous experiments, there was considerable evidence that the majority of subjects preferred the increasing sequence. For the severe headaches choices, 35 (54%) subjects consistently chose the increasing sequence whereas only 12 (18%) subjects consistently chose the decreasing sequence. For the mild coughing choices, 29 (45%) subjects consistently chose the increasing sequence whereas only 12 (18%) subjects consistently chose the decreasing sequence. Lastly, for the severe headaches choices, 59 (91%) subjects consistently chose the increasing sequence whereas only 1 (2%) subject consistently chose the decreasing sequence.

As in Experiment 3, we computed an "independence score" for each test for each subject. Each test was coded as 1 if independence was satisfied twice by that subject, 0 if independence was violated twice by that subject, and 0.5 if the subject did not replicate one or both of their choices. The mean independence scores (and 95% confidence intervals) for the 15 tests are shown in Figure 18. Nine of the 15 intervals are unambiguously above the 0.5 line, indicating reliable satisfaction of independence. For 3 tests (#2, #8, and #13), the lower bound of the confidence interval is very close to 0.5. Finally, for the three tests already described (#6, #7, and #11), the mean itself is close to 0.5.

To determine whether there was a sizable minority of subjects who consistently violated independence, we computed the sum of each subject's independence scores over 15 tests. The sum ranges from 0 (indicating maximum violation of independence) to 15 (indicating maximum satisfaction of independence). A score of 7.5 indicates neither violation nor satisfaction of independence.

The frequency distribution of the sums (Figure 19) does not indicate a clear bimodal distribution. 49 subjects' (75%) sums were greater than 7.5, 4 (6%) were equal to 7.5, and 12 (18%) were less than 7.5. The mean sum was 9.3, and the 95% confidence interval around this mean had a lower bound of 8.6 and an upper bound of 9.9.

Regarding subjects' answers to the strategy questions, subjects generally reported using the strategy that gave poor health sooner than later (strategy X). The percentage of subjects who chose this strategy was 78% for severe headaches, 74% for mild coughing, and 100% for painful procedures.

Finally, we computed three discounting scores for each subject: one for the 20 severe headaches choices, one for the 20 mild coughing choices, and one for the 20 painful procedure choices. Each score ranged from -10 (indicating maximum negative discounting) to 10 (indicating maximum positive discounting). The frequency distributions for these discounting scores are in Figures 20, 21, and 22. These distributions do not appear bimodal, but rather positively skewed. As in previous experiments, however, more subjects consistently used negative discounting than subjects who consistently used positive discounting. The effect was strongest for the painful procedure set: 64 of 65 subjects had a negative discounting score. Almost all subjects preferred to expedite the painful procedures rather than postpone them.

DISCUSSION

Independence was generally satisfied in Experiment 4. There were 12 tests that clearly satisfied independence, and 3 tests that fell on the border between satisfaction and violation. The pattern of violations was not the same for all 3 of these latter tests. One test (#6) indicated no systematic pattern of violation, another (#7) indicated systematic preference ***against*** spreading, and a third (#11) indicated systematic preference *for* spreading. The relative rarity and inconsistency of the violations lead us to conclude that the independence assumption of the discounted QALY model is quite realistic.

We had hypothesized that more violations would occur because the sequences were created to maximize spreading. However, subjects did not display a consistent preference for well-spread sequences over poorly-spread sequences (the only exception was test #11). As in all previous experiments, however, subjects' choices indicating a tendency toward negative discounting rather than positive discounting. This tendency was strongest in the painful procedure set of choices in which all but one subject had a negative discounting score.

Table 3: Independence Tests in Experiment 4

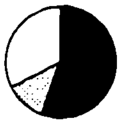
"E" denotes "excellent health"; "S" denotes "severe headaches"; "M" denotes "mild coughing"; "N" denotes "no procedure"; "P" denotes "painful procedure"; "iff" denotes "if and only if".

Years:	1	1	10	1	10	1	1		1	1	10	1	10	1	1
test #1	E	E	E	E	E	E	S	>	E	E	E	S	E	E	E
iff	S	S	S	E	S	S	S	>	S	S	S	S	S	S	E
test #2	E	E	E	S	E	E	E	>	S	E	E	E	E	E	E
iff	E	S	S	S	S	S	S	>	S	S	S	E	S	S	S
test #3	E	E	E	E	E	S	S	>	S	E	E	E	E	E	S
iff	E	S	S	S	S	S	E	>	S	S	S	S	S	E	E
test #4	S	E	E	E	E	E	S	>	S	S	E	E	E	E	E
iff	E	E	S	S	S	S	S	>	E	S	S	S	S	S	E
test #5	E	E	E	E	E	S	S	>	S	S	E	E	E	E	E
iff	E	E	S	S	S	S	S	>	S	S	S	S	S	E	E
test #6	E	E	E	E	E	E	M	>	E	E	E	M	E	E	E
iff	M	M	M	E	M	M	M	>	M	M	M	M	M	M	E
test #7	E	E	E	M	E	E	E	>	M	E	E	E	E	E	E
iff	E	M	M	M	M	M	M	>	M	M	M	E	M	M	M
test #8	E	E	E	E	E	M	M	>	M	E	E	E	E	E	M
iff	E	M	M	M	M	M	E	>	M	M	M	M	M	E	E
test #9	M	E	E	E	E	E	M	>	M	M	E	E	E	E	E
iff	E	E	M	M	M	M	M	>	E	M	M	M	M	M	E
test #10	E	E	E	E	E	M	M	>	M	M	E	E	E	E	E
iff	E	E	M	M	M	M	M	>	M	M	M	M	M	E	E
Days:	1	1	1	1	1	1	1		1	1	1	1	1	1	1
test #11	N	N	N	N	N	N	P	>	N	N	N	P	N	N	N
iff	P	P	P	N	P	P	P	>	P	P	P	P	P	P	N
test #12	N	N	N	P	N	N	N	>	P	N	N	N	N	N	N
iff	N	P	P	P	P	P	P	>	P	P	P	N	P	P	P
test #13	N	N	N	N	N	P	P	>	P	N	N	N	N	N	P
iff	N	P	P	P	P	P	N	>	P	P	P	P	P	N	N

Table 3 (continued)

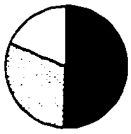
Days:	1	1	1	1	1	1	1		1	1	1	1	1	1	1
test #14	P	N	N	N	N	N	P	>	P	P	N	N	N	N	N
iff	N	N	P	P	P	P	P	>	N	P	P	P	P	P	N
test #15	N	N	N	N	N	P	P	>	P	P	N	N	N	N	N
iff	N	N	P	P	P	P	P	>	P	P	P	P	P	N	N

test #1 A>B iff A'>B' — EEEEEES>EEESEEE iff SSSESSS>SSSSSSE



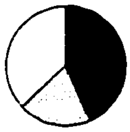
	choose A'	choose B'
ch. A	9	4
ch. B	4	27

test #2 A>B iff A'>B' — EEESEEE>SEEEEEE iff ESSSSSS>SSSESSS



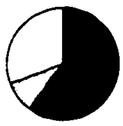
	choose A'	choose B'
ch. A	4	4
ch. B	16	29

test #3 A>B iff A'>B' — EEEEESS>SEEEEES iff ESSSSSE>SSSSSEE



	choose A'	choose B'
ch. A	9	8
ch. B	5	19

test #4 A>B iff A'>B' — SEEEEES>SSEEEEE iff EESSSSS>ESSSSSE



	choose A'	choose B'
ch. A	6	1
ch. B	5	33

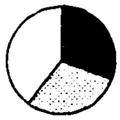
test #5 A>B iff A'>B' — EEEEESS>SSEEEEE iff EESSSSS>SSSSSEE

	choose A'	choose B'
ch. A	12	0
ch. B	6	35

Figure 15: Independence for Severe Headaches Sequences in Experiment 4

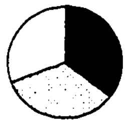
- ■ **subjects who satisfied independence twice**
- □ **subjects who violated independence twice**
- □ **subjects who did not replicate choices**

test #6: A>B EEEEEEM>EEEMEEE
iff
A'>B' MMMEMMM>MMMMMME



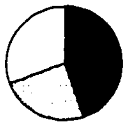
	choose A	choose B
ch. A	3 ■	9 ▒
ch. B	10 ▒	17 ■

test #7: A>B EEEMEEE>MEEEEEE
iff
A'>B' EMMMMMM>MMMEMMM



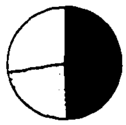
	choose A	choose B
ch. A	3 ■	5 ▒
ch. B	17 ▒	20 ■

test #8: A>B EEEEEMM>MEEEEEM
iff
A'>B' EMMMMME>MMMMMEE



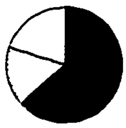
	choose A	choose B
ch. A	8 ■	6 ▒
ch. B	10 ▒	21 ■

test #9: A>B MEEEEEM>MMEEEEE
iff
A'>B' EEMMMMM>EMMMMME



	choose A	choose B
ch. A	3 ■	3 ▒
ch. B	12 ▒	29 ■

test #10: A>B EEEEEMM>MMEEEEE
iff
A'>B' EEMMMMM>MMMMMEE

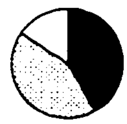


	choose A	choose B
ch. A	12 ■	0 ▒
ch. B	11 ▒	29 ■

Figure 16: Independence for Mild Coughing Sequences in Experiment 4

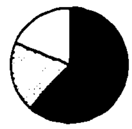
■ **subjects who satisfied independence twice**
□ **subjects who violated independence twice**
□ **subjects who did not replicate choices**

test #11 A>B iff A'>B'
NNNNNNP>NNNPNNN iff PPPNPPP>PPPPPPN



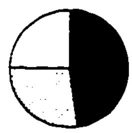
	choose A'	choose B'
ch. A	5	6
ch. B	22	22

test #12 A>B iff A'>B'
NNNPNNN>PNNNNNN iff NPPPPPP>PPPNPPP



	choose A'	choose B'
ch. A	0	2
ch. B	11	40

test #13 A>B iff A'>B'
NNNNNPP>PNNNNNP iff NPPPPPN>PPPPPNN

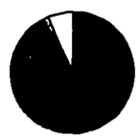


	choose A'	choose B'
ch. A	2	13
ch. B	5	29

test #14 A>B iff A'>B'
PNNNNNP>PPNNNNN iff NNPPPPP>NPPPPPN

	choose A'	choose B'
ch. A	1	7
ch. B	7	45

test #15 A>B iff A'>B'
NNNNNPP>PPNNNNN iff NNPPPPP>PPPPPNN



	choose A'	choose B'
ch. A	1	0
ch. B	1	59

Figure 17: Independence for Painful Procedure Sequences in Experiment 4

■ **subjects who satisfied independence twice**

□ **subjects who violated independence twice**

□ **subjects who did not replicate choices**

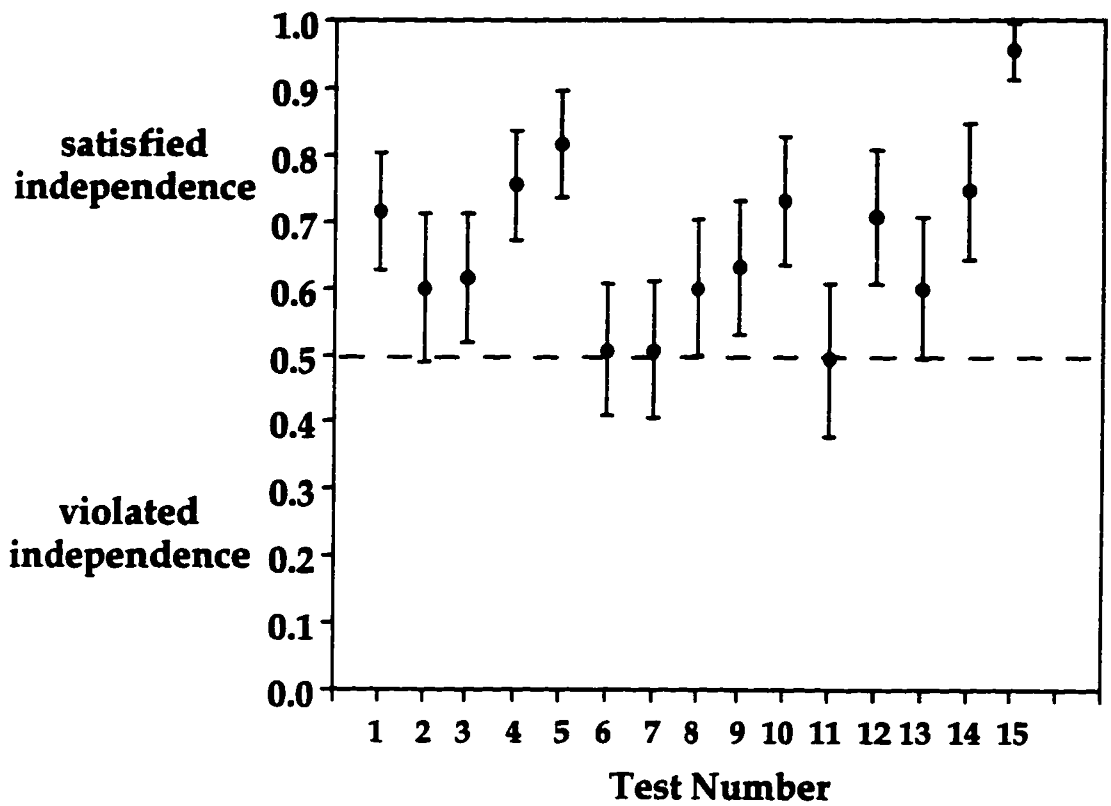


Figure 18: Means and 95% Confidence Intervals for Independence Tests in Experiment 4

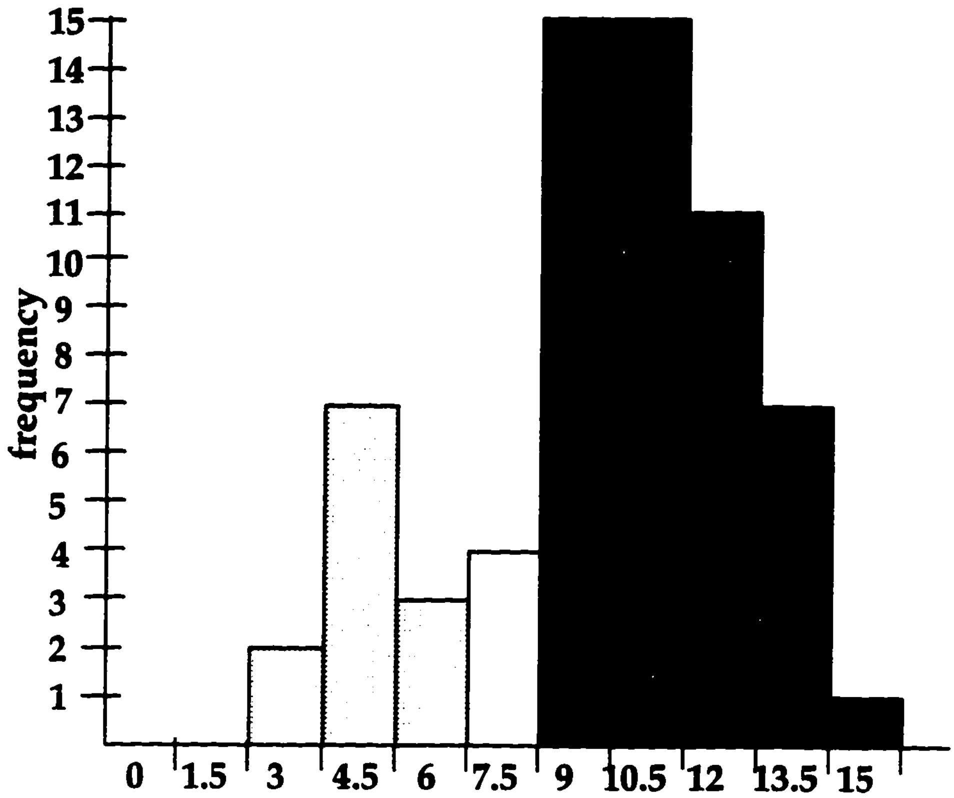


Figure 19: Frequency Distribution of the Independence Score in Experiment 4

Each section of the horizontal axis represents three possible discounting scores. There is one exception to this: the section in the middle (7.5) represents only a score of 7.5.

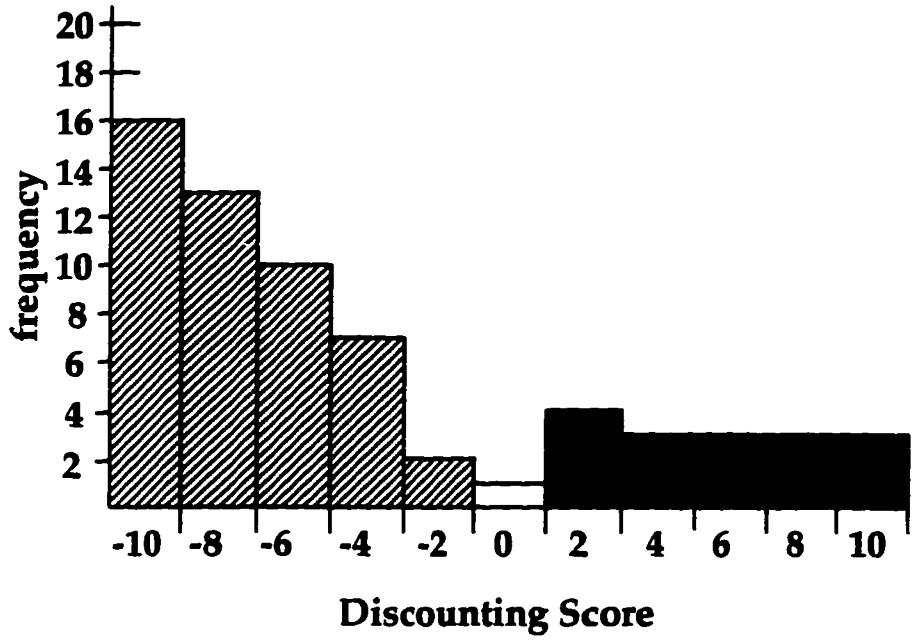


Figure 20: Frequency Distribution of the Discounting Score for Severe Headaches Sequences in Experiment 4

Each section of the horizontal axis represents two possible discounting scores, as in Figure 2. There is one exception to this: the section in the middle (0) represents only a score of 0.

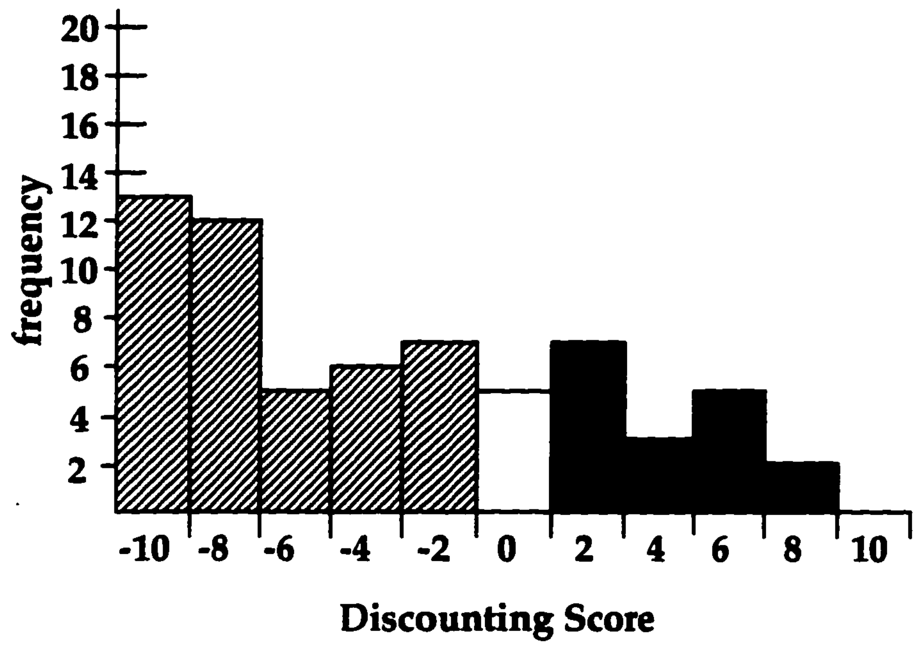


Figure 21: Frequency Distribution of the Discounting Score for Mild Coughing Sequences in Experiment 4

Each section of the horizontal axis represents two possible discounting scores, as in Figure 2. There is one exception to this: the section in the middle (0) represents only a score of 0.

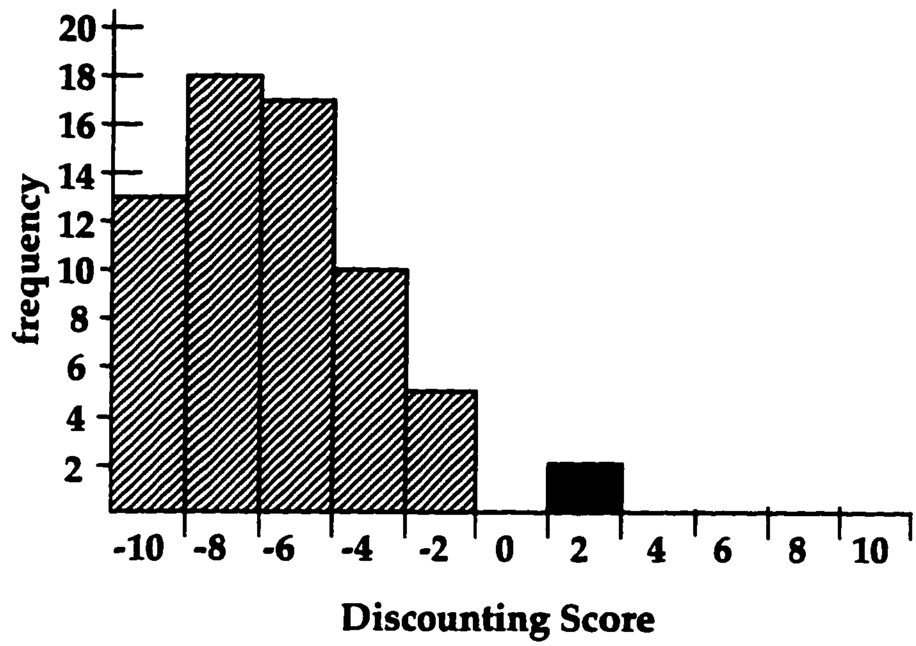


Figure 22: Frequency Distribution of the Discounting Score for Painful Procedure Sequences in Experiment 4

Each section of the horizontal axis represents two possible discounting scores, as in Figure 2. There is one exception to this: the section in the middle (0) represents only a score of 0.

CHAPTER 6: GENERAL DISCUSSION

There were three central findings in this dissertation:

1) When choosing between health sequences, people tend to use negative discounting.

2) The savoring/dread explanation of negative discounting is insufficient.

3) When choosing between health sequences, people tend to satisfy independence.

These findings will now be discussed in detail.

In all four experiments, subjects tended to prefer sequences in which the poorer health states were experienced sooner rather than later. Expediting the poorer health states, therefore, was more appealing than postponing them. If the discounted QALY model is true, then this preferences implies a negative discount rate.

This finding contradicts the assumption of positive discounting that is made in virtually all applications of the discounted QALY model. One explanation for the contradiction lies in the distinction between preferences for sequences and preferences for single outcomes. Preferences for single outcomes in the health domain typically suggest positive discounting (refer again to pages 17-20). For sequences, on the other hand, the general tendency is negative discounting. Therefore, if an application of the discounted QALY model only claims to be relevant for single outcomes, then the assumption of positive discounting is realistic. If, however, the application is also intended for health sequences, then the assumption of positive discounting should be questioned.

In Experiment 2, we empirically tested the savoring/dread explanation of negative discounting. According to this explanation, subjects choose increasing sequences over decreasing sequences in order to maximize savoring and minimize dread. Our manipulation of foreknowledge did not support this explanation. Preference patterns were the same regardless of whether foreknowledge was eliminated, emphasized, or not mentioned. One possible but unlikely reason for this result is that our manipulation was too weak to reveal an effect. Other explanations, such as adaptation/loss aversion and recency, deserve consideration to understand why people appear to use negative discounting.

In Experiments 3 and 4, subjects reliably satisfied independence in 39 of 42 tests. For the other three tests, independence was neither satisfied nor violated, and there was not a single consistent pattern of violation. We concluded that independence is generally satisfied: preferences do not switch if tied intervals are altered. This conclusion is equivalent to stating that there is no "sequence effect" as suggested by Krabbe and Bonsel (in press). A sequence's overall valuation can be accurately estimated by adding the valuations of its separate intervals.

Based on these findings, we suggest that a parsimonious and reasonably accurate model of health sequence preferences would be a negatively discounted QALY model. Formally, the model would be written as follows:

$$\text{Discounted } QALYs(\text{1st year in health state } Q_1\text{, then 2nd year in health state } Q_2\text{, then } \ldots \text{ Yth year in health state } Q_Y) = \sum_{i=1}^{Y} \frac{H(Q_i)}{(1+d)^{i-1}} \text{ where } d < 0 \qquad \text{Equation 13}$$

This equation is identical to Equation 4 (page 5) except that here d is required to be negative. For the majority of subjects, this requirement is realistic.

How would this model be used to assign a number to a health sequence that accurately reflects its overall value to a particular subject? The process would have three steps. First, the analyst would determine that subject's H(Q) for each of the discrete health states in the sequence. Second, the analyst would estimate what negative discount rate d the subject uses when choosing between health sequences. Third, the analyst would compute the negatively-discounted QALYs for each interval separately and then add the intervals to produce the overall number of QALYs of the sequence.

One limitation of the conclusions in this thesis is that all subjects were young healthy college students. They probably had little experience with serious illnesses or making difficult medical decisions. Therefore, it is important to investigate discounting and independence for health sequences among actual patients. We are currently conducting a study at the Seattle Department of Veterans' Affairs Medical Center with patients who have hepatitis-C. In a self-administered questionnaire, patients are making choices between medically realistic health sequences. From the results, we will be able to test independence as well as infer the valence of discounting. The patients' responses may provide more evidence for negative discounting and satisfaction of independence for health sequences. If so, we would have some assurance that the findings from Experiments 1-4 are not unique to college students but rather reflect general preference patterns.

BIBLIOGRAPHY

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. Econometrica, 21, 503-546.

Cairns, J.A. (1992). Future discounting: Health, wealth, and time preference. Project Appraisal, 7(1), 31-40.

Chapman, G.C. (1996a). Temporal discounting and utility for health and money. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(3), 771-791.

Chapman, G.C. (1996b). Expectations and preferences for sequences of health and money. Organizational Behavior and Human Decision Processes, 67(1), 59-75.

Chapman, G.B. & Elstein, A.S. (1995). Valuing the future: Temporal discounting of health and money. Medical Decision Making, 15(4), 373-386.

Christensen-Szalanski, J.J.J. (1984). Discount functions and the measurement of patient's values: Women's decisions during childbirth. Medical Decision Making, 4 (1), 47-57.

Coombs, C.H., Dawes, R.M. & Tversky, A. (1970). Mathematical Psychology: An Elementary Introduction. Prentice Hall: New Jersey.

Coyle, D. & Tolley, K. (1992). Discounting of health benefits in the pharmacoeconomic analysis of drug therapies: An issue for debate? Pharmacoeconomics, 2(2), 153-162.

Drummond, M.F., Stoddart, G.L., & Torrance, G.W. (1987). Methods for the Economic Evaluation of Health Care Programmes. Oxford: Oxford University Press.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. Quarterly Journal of Economics, 75, 643-669.

Froberg, D.G. & Kane, R.L. (1989). Methodology for measuring health-state preferences-II: Scaling methods. Journal of Clinical Epidemiology, 42(5), 459-471.

Fuchs, V.R. (1982). Time preference and health: An exploratory study. In V.R. Fuchs (Ed.), Economic Aspects of Health. Chicago: University of Chicago Press.

Ganiats, T.G. (1994). Discounting in cost-effectiveness research. Medical Decision Making, 14(3), 298-300.

Goldstein, W.M. & Einhorn, H.J. (1987). Expression theory and the preference reversal phenomenon. Psychological Review, 94, 236-254.

Goodin, R.E. (1982). Discounting discounting. Journal of Public Policy, 2, 53-72.

Helson, H. (1964). Adaptation-level theory: An experimental and systematic approach to behavior. New York: Harper & Row.

Jevons, W.S. (1871). Theory of Political Economy. London: Macmillan.

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.

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47, 363-391.

Keeler, E.B. & Cretin, S. (1983). Discounting of life-saving and other nonmonetary effects. Management Science, 29, 300-306.

Krabbe, P.F.M. & Bonsel, G.J. (in press). Sequence effects, health profiles, and the QALY model. Medical Decision Making.

Linville, P. & Fischer, G. (1991). Preferences for combining or separating events: A social application of prospect theory and mental accounting. Journal of Personality and Social Psychology, 60, 5-23.

Lipscomb, J. (1989). Time preference for health in cost-effectiveness analysis. Medical Care, 27(3), S233-S253.

Loewenstein, G. (1987). Anticipation and the valuation of delayed consumption. Economic Journal, 97, 666-684.

Loewenstein, G. (1992). The fall and rise of psychological explanations in the economics of intertemporal choice. In G. Loewenstein and J. Elster (Eds.), Choice Over Time. Russell Sage Foundation: New York.

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

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

Loewenstein, G. & Sicherman, N. (1991). Do workers prefer increasing wage profiles? Journal of Labor Economics, 9, 67-84.

MacKeigan, L.D., Larson, L.N., Draugalis, J.R., Bootman, J.L. & Burns, L.R. (1993). Time preference for health gains versus health losses. PharmacoEconomics, 3(5), 374-386.

McNeil, B.J., Weichelsbaum, R., & Pauker, S.G. (1981). Tradeoffs between quality and quantity of life in laryngeal cancer. New England Journal of Medicine, 305, 982-987.

Parsonage, M. & Neuberger, H. (1992). Discounting and health benefits. Health Economics, 1(1), 71-76.

Pliskin, J.S., Shepard, D.S., & Weinstein, M.C. (1980). Utility functions for life years and health status. Operations Research, 28, 206-224.

Rae, J. (1834, reprinted 1905). The Sociological Theory of Capital. London: Macmillan.

Redelmeier, D.A. & Heller, D.N. (1993). Time preference in medical decision making and cost-effectiveness analysis. Medical Decision Making, 13(3), 212-217.

Redelmeier, D.A., Heller, D.N., & Weinstein, M.C. (1994). Time preference in medical economics: Science or religion? Medical Decision Making, 14(3), 301-303.

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.

Ross, W.T. & Simonson, I. (1991). Evaluating pairs of experiences: A preference for happy endings. Journal of Behavioral Decision Making, 4, 273-282.

Roth, C.A., Ing, R.T., & Ross, D.A. (1978). letter to New England Journal of Medicine, 298(19), 1088.

Rose, D.N. & Weeks, M.G. (1988). Individuals' discounting of future monetary gains and health states. Medical Decision Making, 18(4), 334.

Senior, N.W. (1836). An Outline of the Science of Political Economy. London: Clowes and Sons.

Thaler, R. & Johnson, E. (1990). Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. Management Science, 36, 643-660.

Tversky, A. & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. Quarterly Journal of Economics, 106, 1039-1061.

Tversky, A., Slovic, P. & Kahneman, D. (1990). The causes of preference reversal. American Economic Review, 80, 204-217.

Varey, C. & Kahneman, D. (1992). Experiences extended across time: Evaluation of moments and episodes. Journal of Behavioral Decision Making, 5, 169-185.

von Neumann, J. & Morganstern, O. (1947). Theory of Games and Economic Behavior. 2nd Ed. Princeton University Press: Princeton, NJ.

Weinstein, M.C. (1993). Time preference studies in the health care context. Medical Decision Making, 13(3), 218-219.

Weinstein, M.C., Fineberg, H.V. & Elstein, A.S. (1980). Clinical Decision Analysis. Philadelphia: W.B. Saunders.

Weinstein, M.C. & Stason, W.B. (1977). Foundations in cost-effectiveness analysis for health and medical practices. New England Journal of Medicine, 296, 716-721.

Williams, A. (1981). Welfare economics and health status measurement. In J. van der Gaag and M. Perlman (Eds.), Health, Economics, and Health Economics. North-Holland.

APPENDIX A: HEALTH STATE DESCRIPTIONS

Persistent coughing. This means that, once a week for a 10-year period, you will have a coughing fit (you will cough repeatedly). This coughing will be very annoying, and it will last about 1 hour. Neither cough syrup nor any other kind of medication will relieve the coughing; you'll just have endure it.

Mild headaches. This means that, once a week for a 10-year period, you will have a mild headache. Each headache will last about 1 hour. The pain will not be severe, but it will be somewhat bothersome. Neither aspirin nor any other kind of medication will relieve these headaches; you'll just have to endure them.

Excellent health. This means that you will have no health problems at all for a 10-year period.

APPENDIX B: PROOF THAT DISCOUNTING IMPLIES PREFERENCES

This appendix contains a proof that positive discounting implies that (moderate, excellent, poor) should be preferred to (poor, excellent, moderate).

Notation:

w_1=weight for interval 1

w_2=weight for interval 2

w_3=weight for interval 3

U(E)=utility of excellent health

U(M)=utility of moderate health

U(P)=utility of poor health

According to positive discounting, $w_1>w_2>w_3$. Choose constants $x>0$ and $y>0$ such that:

$$w_1 = w_2 + x$$

$$w_3 = w_2 - y$$

Assuming a monotonic utility function, U(E)>U(M)>U(P). Choose constants A>0 and B>0 such that:

$$U(E) = U(M) + A$$

$U(P) = U(M) - B$

According to the QALY formula,

$$U(\text{moderate, excellent, poor}) = w_1U(M) + w_2U(E) + w_3U(P)$$

$$U(\text{moderate, excellent, poor}) = (w_2+x)U(M) + w_2(U(M)+A) + (w_2-y)(U(M)-B)$$

$$U(\text{moderate, excellent, poor}) = 3w_2U(M) + xU(M) + w_2A - w_2B - yU(M) + yB$$

Also, according to the QALY formula,

$$U(\text{poor, excellent, moderate}) = w_1U(P) + w_2U(E) + w_3U(M)$$

$$U(\text{poor, excellent, moderate}) = (w_2+x)(U(M)-B) + w_2(U(M)+A) + (w_2-y)U(M)$$

$$U(\text{poor, excellent, moderate}) = 3w_2U(M) - w_2B + xU(M) - xB + w_2A - yU(M)$$

The two equations share five terms which can be cancelled: $3w_2U(M)$, $xU(M)$, w_2A, $-w_2B$, and $-yU(M)$. That leaves only yB in the first sequence and $-xB$ in the second sequence. Because x, y, and B are all positive, the first term is positive but the second term is negative. Therefore, U(moderate, excellent, poor) > U(poor, excellent, moderate).

A proof can be given from the perspective of negative discounting which implies the opposite preference.

APPENDIX C: COUNTEREXAMPLE

This appendix contains a counterexample to show that positive discounting does not necessarily imply a particular preference between (moderate, poor, excellent) and (poor, excellent, moderate).

Suppose that (consistent with positive discounting):

$w_1 = 0.7$

$w_2 = 0.2$

$w_3 = 0.1$

Also suppose that (consistent with a monotonic utility function):

U(E)=1

U(M)=0.9

U(P)=0.1

Then, according to the QALY formula,

U(moderate, poor, excellent)= $(0.7)(0.9) + (0.2)(0.1) + (0.1)(1) = 0.75$

U(poor, excellent, moderate)= $(0.7)(0.1) + (0.2)(1) + (0.1)(0.9) = 0.36$

Therefore, using the above weights and utilities, the sequence (moderate, poor, excellent) would be preferred.

However, suppose the following weights and utilities were used:

$w_1 = 0.5$

$w_2 = 0.4$

$w_3 = 0.1$

U(E)=1

U(M)=0.2

U(P)=0.1

Then, according to the QALY formula,

U(moderate, poor, excellent)= $(0.5)(0.2) + (0.4)(0.1) + (0.1)(1) = 0.24$

U(poor, excellent, moderate)= $(0.5)(0.1) + (0.4)(1) + (0.1)(0.2) = 0.47$

Those weights and utilities suggest the opposite preference.

Therefore, the preference depends on the choice of weights and utilities. So positive discounting does not imply a particular preference between these two sequences. A similar proof can be given from the perspective of negative discounting.

APPENDIX D: GENERAL FOREKNOWLEDGE INSTRUCTIONS

The foreknowledge instructions for the N group in Experiment 2 at the beginning of the experiment are below.

For the purposes of this experiment, it is important to remember this: BOTH OPTIONS INVOLVE SURPRISE EVENTS.

In other words, you will be asked to choose between different events whose occurrence you will not anticipate.

For example, would you rather have a surprise encounter and conversation with UW President Richard McCormick, or a surprise encounter and conversation with Governor Gary Locke?

Notice that, because these are surprise encounters, you would not be able to prepare for them. So you could not prepare questions that you would want to ask, nor bring your friends with you, nor do anything else that you would do if you knew about the encounter in advance.

The foreknowledge instructions at the beginning of the experiment for the F group are below.

For the purposes of this experiment, it is important to remember this: BOTH OPTIONS INVOLVE ANTICIPATED EVENTS.

In other words, you will be asked to choose between different events whose occurrence you will anticipate.

For example, would you rather have a conversation with UW President Richard McCormick, or a conversation with Governor Gary Locke? Assume that both conversations are planned in advance.

Notice that, because these are anticipated events, you would be able to prepare for them. So you could prepare questions that you would want to ask, or bring your friends with you, or do anything else that you would do if you knew about the conversation in advance.

APPENDIX E: RESTAURANT QUESTIONS

1) Which would you prefer if both were free?

a: Dinner at a fancy French restaurant

b: Dinner at a local Greek restaurant

2) Which would you prefer?

a: Dinner at the French restaurant in 1 month

b: Dinner at the French restaurant in 2 months

3) Which would you prefer?

a: Dinner at the Greek restaurant in 1 month

b: Dinner at the Greek restaurant in 2 months

4) Which would you prefer?

a: Dinner at the French restaurant in 1 month, and dinner at the Greek restaurant in 2 months

b: Dinner at the Greek restaurant in 1 month, and dinner at the French restaurant in 2 months

APPENDIX F: ENTERTAINMENT QUESTIONS

1) Which would you prefer if both were free?

a: Going to a popular theatrical play

b: Going to a popular movie

2) Which would you prefer?

a: Going to the play in 1 month

b: Going to the play in 2 months

3) Which would you prefer?

a: Going to the movie in 1 month

b: Going to the movie in 2 months

4) Which would you prefer?

a: Going to the play in 1 month, and going to the movie in 2 months

b: Going to the movie in 1 month, and going to the play in 2 months

APPENDIX G: FOREKNOWLEDGE INSTRUCTIONS FOR RESTAURANT AND ENTERTAINMENT QUESTIONS

The foreknowledge instructions for the N group in Experiment 2 prior to the restaurant/entertainment questions are below. (changes for the entertainment questions are in brackets)

Suppose that it is Friday at 6:00 p.m, and you have not yet eaten dinner [do not have plans for the evening]. Suddenly, a friend calls and says that he/she has won two free dinners [tickets] at a good restaurant [show] for that same evening. Your friend invites you to go to dinner [the show], completely free of charge.

Notice that, because the dinner [show] will be that very same evening, you would not be able to plan in advance for it. Instead, it would be a surprise event.

The foreknowledge instructions prior to the restaurant/entertainment questions for the F group are below. (changes for the entertainment questions are in brackets)

Suppose that a friend calls and says that he/she has won two free dinners [tickets] at a good restaurant [show]. Your friend invites you to go to dinner [the show], completely free of charge. The two of you can have the dinner [go to the show] at any time in the future.

Notice that, because the dinner [show] would be in the future, you would be able to plan in advance for it. Therefore, it would be a anticipated event.

APPENDIX H: FOREKNOWLEDGE INSTRUCTIONS FOR THE TIME TRADEOFF TASK

The foreknowledge instructions for the N group in Experiment 2 prior to the "severe headaches - 10 years" time tradeoff tasks are below. (changes for persistent coughing and/or 2 year time period are in brackets)

Suppose, in addition to the severe headaches [persistent coughing], you only have 10 [2] years of life remaining.

If you knew this fact in advance, you might change your lifestyle because you only had 10 [2] years of life remaining. For example, you might travel to places you've always wanted to go, devote more time to your favorite hobby, or spend more time with people you love.

For the purposes of this experiment, however, we want you to assume that YOU WOULD NOT KNOW IN ADVANCE that you only have 10 [2] years of life remaining. So you will live your remaining years UNAWARE of how much longer you have to live. In other words, you will not know your own time of death.

Because you are unaware that you have 10 [2] years to live, you would NOT change your lifestyle. Instead, you would live your life exactly the way you do now.

Think about what it would be like to live with severe headaches [persistent coughing], and to live for only 10 [2] more years.

What if you could remove all of the with severe headaches [persistent coughing] (and therefore live in excellent health), but it would mean that you would not live for the full 10 [2] years?

Think about how much time you would be willing to give up in order to relieve yourself of the severe headaches [persistent coughing].

Remember, however, that under no circumstances would you know for sure exactly how many years of life you have left.

Therefore, any amount of life in excellent health would be lived without the knowledge of the time of your death.

Also, any amount of life with severe headaches [persistent coughing] would be lived without the knowledge of the time of your death.

The foreknowledge instructions for the F group prior to the "severe headaches - 10 years" time tradeoff tasks are below. (changes for persistent coughing and/or 2 year time period are in brackets)

Suppose, in addition to the severe headaches [persistent coughing], you only have 10 [2] years of life remaining.

If you knew this fact in advance, you might change your lifestyle because you only had 10 [2] years of life remaining. For example, you might travel to places you've always wanted to go, devote more time to your favorite hobby, or spend more time with people you love.

For the purposes of this experiment, we want you to assume that YOU WOULD KNOW IN ADVANCE that you only have 10 [2] years of life *remaining. So you will live your remaining years AWARE of how much longer you have to live. In other words, you will know your own time of death.*

Because you are aware that you have 10 [2] years to live, you might change your lifestyle in the above ways.

Think about what it would be like to live with severe headaches [persistent coughing], and to live for only 10 [2] more years.

What if you could remove all of the with severe headaches [persistent coughing] (and therefore live in excellent health), but it would mean that you would not live for the full 10 [2] years?

Think about how much time you would be willing to give up in order to relieve yourself of the severe headaches [persistent coughing].

Remember, however, that you would always know for sure exactly how many years of life you have left.

Therefore, any amount of life in excellent health would be lived with the knowledge of the time of your death.

Also, any amount of life with severe headaches [persistent coughing] would be lived with the knowledge of the time of your death.

APPENDIX I: FOREKNOWLEDGE INSTRUCTIONS FOR HEALTH SEQUENCES

The foreknowledge instructions for the N group in Experiment 2 prior to the health sequence questions are below.

We want you to assume that you would never know how many years of remaining life you have. Therefore, you could not change your lifestyle so that you would "live life to the fullest" in your 30 years of remaining life. Instead, your lifestyle would be exactly the same as it is now.

Also, you would never know about the changes in your health that might occur in the future. Therefore, if the health sequence is:

You live for the first 10 years in excellent health

You live for the next 10 years with persistent coughing

You live for the last 10 years with severe headaches

assume that during the first 10 years in excellent health, you would not know that your health will later change to "persistent coughing", nor would you know that your health will change even later to "severe headaches".

The foreknowledge instructions for the F group in Experiment 4 prior to the health sequence questions are below.

We want you to assume that you would always know how many years of remaining life you have. Therefore, you could change your lifestyle so that you would "live life to the fullest" in your 30 years of remaining life.

Also, you would always know about the changes in your health that might occur in the future. Therefore, if the health sequence is:

You live for the first 10 years in excellent health

You live for the next 10 years with persistent coughing

You live for the last 10 years with severe headaches

assume that during the first 10 years in excellent health, you would know that your health will later change to "persistent coughing", and also you would know that your health will change even later to "severe headaches".

APPENDIX J: ASSUMPTION CHECK

Which of the following statements best describes the assumption you made when answering these questions?

a: In this experiment, I assumed that I would know in advance about future outcomes that might happen.

b: In this experiment, I assumed that I would not know in advance about future outcomes that might happen.

APPENDIX K: ADDITIONAL HEALTH STATE DESCRIPTIONS

Severe headaches. This means that, once every day for a 10-year period, you will have a severe headache. The pain will be severe and very bothersome. Each headache will last about 1 hour. Neither aspirin nor any other kind of medication will relieve these headaches; you'll just have to endure them.

Mild coughing. This means that, once every day for a 10-year period, you will have a mild coughing spell (you will cough a lot). It will not be very painful, but it will be slightly bothersome. The coughing will last about 1 hour. Neither cough syrup nor any other kind of medication will relieve the coughing; you'll just have endure it.

Excellent health. This means that you will have no health problems at all for a 10-year period.

APPENDIX L: DESCRIPTION OF HYPOTHETICAL PAINFUL PROCEDURE

Suppose that you have a very serious but TREATABLE illness. That is, your doctor can completely cure the illness. Therefore, after you have been treated, your life will return to normal. In order to cure the illness, however, you will have to undergo a series of painful medical procedures. These procedures will be very unpleasant and painful, but they are absolutely necessary.

Although each procedure would be painful, you would feel pain ONLY on the day of the procedure. By the next day, you would feel perfectly fine. Also, you would never have more than one procedure on the same day. Remember: the pain does not carry over into the next day. The pain is only felt on the actual day of the procedure.

An example of a painful medical procedure is a "spinal tap". In this procedure, the doctor inserts a needle into the spine to obtain a small sample of spinal fluid. This procedure is used to diagnose meningitis, an inflammation of the membranes covering the brain and spinal cord. A spinal tap is painful while it is happening, but when the procedure is over, the patient experiences little or no discomfort.

Suppose that your doctor must perform the procedures in the upcoming week. On some of the days, there will be a painful medical procedure. On the other days, there will be no procedure and you feel perfectly fine.

In the questions that follow, you will be presented with two sequences: Sequence "a" and Sequence "b". Each sequence involves a SEVEN

DAY period in which you would undergo a painful medical procedure on some days, and you would not undergo any procedure on other days.

APPENDIX M: STRATEGY QUESTIONS

1) Think back to the questions that involved SEVERE HEADACHES. When you were answering those questions, which of the two strategies below best describes your approach?

X: I wanted to get the severe headaches over with as soon as possible. Then, I could live in excellent health without having to worry about it.

Y: I wanted to postpone the severe headaches into the future as much as possible. That way, I could live in excellent health while I'm still young.

2) Think back to the questions that involved MILD COUGHING. When you were answering those questions, which of the two strategies below best describes your approach?

X: I wanted to get the mild coughing over with as soon as possible. Then, I could live in excellent health without having to worry about it.

Y: I wanted to postpone the mild coughing into the future as much as possible. That way, I could live in excellent health while I'm still young.

3) Think back to the questions that involved a PAINFUL PROCEDURE. When you were answering those questions, which of the two strategies below best describes your approach?

X: I wanted to get the painful procedures over with as soon as possible. Then, I could resume my normal life without having to worry about them.

Y: I wanted to postpone the painful procedures into the future as much as possible. That way, I would have time to mentally prepare myself for them.

VITA

Jonathan R. Treadwell

University of Washington

1997

Date of birth:	May 4th, 1968
Mailing address:	Department of Psychology, Box 351525 University of Washington Seattle, WA 98195-1525
Phone numbers:	Office: (206) 543-5656 Home: (206) 860-1860
Electronic mail:	treadw@u.washington.edu
Home page:	http://weber.u.washington.edu/~treadw

Education

Bowdoin College, Brunswick, Maine

Major:	Psychology
Minor:	Mathematics
Degree:	B.A., Magna Cum Laude, 1990
Honors:	Dean's List, 1986-1990 Graduated with High Honors in Psychology
Thesis:	"Initial Instance and Hypothesis Generation in Wason's 2-4-6 Task" (advisor: Dr. Melinda Y. Small, Ph.D.)

University of Washington, Seattle, Washington

Major:	Cognition and Perception
Minor:	Quantitative Psychology
Breadths:	Neuroscience, Philosophy
Degree:	Ph.D. expected Spring 1997
Thesis:	Discounting and Independence in Preference Between Health Sequences (advisor: Dr. John M. Miyamoto, Ph.D.)

Professional Positions

1990-1992

Institution: Boston Department of Veterans' Affairs Medical Center, Memory Disorders Research Center, Boston, Massachusetts
Position: Research Assistant
Supervisors: Laird S. Cermak, Ph.D., Mieke Verfaellie, Ph.D.
Duties: Tested patients and controls on neuropsychological experiments via computer. Patients included Korsakoff's amnesics, closed-head-injury patients, encephalitic patients, aneurysm patients.

1990-1992

Institution: Massachusetts General Hospital, Department of Neurology, Boston, Massachusetts
Position: Research Assistant
Supervisors: Walter Koroshetz, M.D., Daniel Willingham, Ph.D.
Duties: Tested Huntington's chorea patients and controls on motor control experiments via computer.

1992-1996

Institution: Department of Psychology, University of Washington, Seattle, Washington
Position: Course Instructor/Teaching Assistant
Supervisors: several, see below
Duties: several, see below

Summer 1994 - Winter 1995

Institution: Department of Family Medicine, University of Washington, Seattle, Washington
Position: Statistical Assistant
Supervisors: Barry Saver, M.D., M.P.H., Thomas R. Taylor, M.D., Ph.D.
Duties: Performed statistical analyses (using SPSS) for a large research project on mammography screening practice in the Puget Sound region. Data collected from over 14,000 women and over 100 physicians.

Summer 1996

Institution: Seattle Veterans' Affairs Medical Center, Seattle, Washington
Position: Patient interviewer
Supervisors: John Spertus, M.D., Cindy Dougherty, Ph.D.

Duties: Assessed utility of angina in cardiovascular disease patients in 100 personal interviews. Utility instruments used: time tradeoff, standard gamble, certainty equivalents.

Courses Taught

Autumn 1993, Spring 1995, Autumn 1995

Course: Laboratory in Human Performance
Description: Students are taught research skills in five standard paradigms in cognitive psychology.

Winter 1996

Course: Probability and Statistics
Description: classical statistics, first quarter

Spring 1996

Course: Analysis of Data
Description: classical statistics, second quarter

Teaching Assistantships

Autumn 1992

Course: Fundamentals of Psychological Research
Professor: Beth Kerr, Ph.D.

Winter 1993, Winter 1994

Course: Probability and Statistics
Professor: Geoffrey R. Loftus, Ph.D.

Spring 1993, Spring 1994

Course: Analysis of Data
Professor: Geoffrey R. Loftus, Ph.D.

Summer 1995

Course: Elementary Psychological Statistics
Professor: John M. Miyamoto, Ph.D.

Membership in Professional Societies

Society for Medical Decision Making 1994-present
American Psychological Society 1993-present

Editorial Activities

Consulting editor, Cognition, February, 1994

Graduate student representative, Guthrie Prize Committee, 1995

Funding Applications

National Science Foundation Graduate Fellowship, 1992, Honorable Mention
Department of Defense Graduate Fellowship, 1993, Honorable Mention
National Science Foundation Graduate Fellowship, 1993, Honorable Mention
National Research Servive Award, 1997, pending

Research Interests

Utility assessment and the incorporation of patient preferences in medical decision making
Preferences for health profiles
Development of computerized utility elicitation techniques
Medical applications of rank dependent utility theory
Measuring the appropriateness of confidence in prior decisions
Neural networks for medical diagnosis

Conference Presentations

Treadwell, J.R., Miyamoto, J.M., & Gonzalez, R. (1995). Models of the utility of life duration. Presented at the 1995 meeting of the Society for Medical Decision Making (Tempe, AZ).

Treadwell, J.R. (1990). Initial instance and hypothesis generation in Wason's 2-4-6 task. Presented at the 1990 meeting of the Maine Psychological Association (Lewiston, ME).

Publications

Treadwell, J.R., & Nelson, T.O. (1996). Availability of information and the aggregation of confidence in prior decisions. Organizational Behavior and Human Decision Processes, 68(1), 13-27.

Treadwell, J.R., & Miyamoto, J.M. (1996). A classic revisited. Journal of Behavioral Decision Making, 9(3), 225-227.

Treadwell, J.R. (1995). Brief note on R.T. Hurlburt's Comprehending Behavioral Statistics. Contemporary Psychology, 40(6), 601.

Schacter, D.L., Church, B., & Treadwell, J.R. (1994). Implicit memory in amnesic patients: Evidence for spared auditory priming. Psychological Science, 5(1), 20-25.

Willingham, D.B., Koroshetz, W.J., Treadwell, J.R., & Bennett, J.P. (1994). Comparison of Huntington's and Parkinson's Disease patients use of advance information. Neuropsychology, 9(1), 39-46.

Schacter, D.L., Cooper, L.A., & Treadwell, J.R. (1993). Preserved priming of novel objects across size transformations in amnesic patients. Psychological Science, 4(5), 331-335.

Verfaellie, M. & Treadwell, J.R. (1993). Status of recognition memory in amnesia. Neuropsychology, 7(1), 5-13.

Under Review

Saver, B., Taylor, T.R, Treadwell, J.R., & Cole, W. Do physicians do as they say? The case of mammography. Submitted to Journal of the American Medical Association.

Projects Underway

Models of the utility of survival duration. (with J. Miyamoto and R. Gonzalez)

On considering the preferences of breast cancer survivors for hormone replacement therapy. (with M. O'Connell)

Converting the Seattle angina questionnaire into a utility measure. (with J. Spertus, J. Miyamoto, H. Starks, and C. Dougherty)

The judgment-of-learning effect in amnesia. (with T. Nelson, J. Grafman, and L. Cermak)

References

John M. Miyamoto, Ph.D.
Department of Psychology, Box 351525
University of Washington
Seattle, Washington 98195-1525
(206) 543-0805
Electronic mail: jmiyamot@u.washington.edu

Geoffrey R. Loftus, Ph.D.
Department of Psychology, Box 351525
University of Washington
Seattle, WA 98195-1525
Office: (206) 543-8874

Fax: (206) 685-3157
Electronic mail: gloftus@u.washington.edu

John Spertus, M.D., M.P.H.
Section of Cardiology
University of Missouri Kansas City
Truman Medical Center
2301 Holmes Street
Kansas City, Missouri 64108
Office: (816) 235-1947
Electronic mail: jspertus@cctr.umkc.edu

Barry Saver, M.D., M.P.H.
Department of Family Medicine, Box 355304
University of Washington
3747 15th Avenue Northeast
Seattle, WA 98195-5304
Office: (206) 543-2461
Electronic mail: barry_saver@fammed.washington.edu

Thomas Taylor, M.D., Ph.D.
Department of Family Medicine, Box 355304
University of Washington
3747 15th Avenue Northeast
Seattle, WA 98195-5304
Office: (206) 543-2461
Electronic mail: tom_taylor@fammed.washington.edu