How to measure discount rates? An experimental comparison of three methods.

David J. Hardisty[1,2], Katherine J. Thompson[2], David H. Krantz[2], & Elke U. Weber[2]

Stanford University[1]

Columbia University[2]

Author note: Please address correspondence to David Hardisty, Knight Management Center, Stanford University, 655 Knight Way, Stanford, CA 94305-7298, dhardisty[at]stanford[dot]edu, 650-498-4497. Support for this research was provided by National Science Foundation grant SES-0820496.

Abstract

Time preferences for financial and air quality gains and losses at delays of up to 50 years were elicited using three different methods: matching, fixed-sequence choice titration, and a dynamic "multiple staircase" choice method. Results indicate that the choice-based methods are prone to influencing participants' discount rates through the magnitude and order of options presented to participants. However, choice-based methods are easier for participants to understand than matching is, and are better at predicting consequential intertemporal choices such as smoking. No consistent advantages were found for the multiple staircase over simple titration. Implications for best practice are discussed.

How to measure discount rates? An experimental comparison of three methods.

Throughout life, people are continually making choices about what to do or have immediately, and what to put off until later. Whenever someone chooses an immediate benefit at the expense of a larger delayed benefit, that person is said to exhibit *temporal discounting* (Samuelson, 1937). Similarly, if someone chooses to avoid an immediate loss in favor of a larger, later loss (for example, postponing a credit-card payment), this represents discounting as well. Laboratory measures of discounting can predict many important real-world behaviors, including credit card debt, smoking, exercise, body-mass index, and infidelity (Chabris, Laibson, Morris, Schuldt, & Taubinsky, 2008; Meier & Sprenger, 2010; Reimers, Maylor, Stewart, & Chater, 2009).

Despite the growing popularity of research on temporal discounting, there is relatively little consensus or empirical research on measurement techniques and best practices. A comprehensive review paper on discounting (Frederick, Loewenstein, & O'Donoghue, 2002) noted a huge heterogeneity in discount rates between studies, and hypothesized that variability in measurement methods might have been a major cause. Fifty-two percent of studies used *choice-based* measures, 31% used *matching*, and 17% used another method.

Measuring discount rates: Choice versus matching

Choice-based methods generally present participants with a series of binary comparisons and use these to infer an indifference point, which is then converted into a discount rate. For example, suppose a participant is presented with a choice between receiving $10 immediately or $11 in one year, and he chooses the immediate option, and that subsequently the participant must decide between $10 or $12 in one year, and he chooses the future option. This pattern of choices

implies that the participant would be indifferent between $10 today and roughly $11.50 in one year. This indifference point can then be converted into a discount rate using a number of different models (discussed below). For example, using the normative, continuously-compounded exponential model, this would yield a discount rate of 14%. The matching method, in contrast, asks for the indifference point directly. For example, it might ask the participant what amount "X" would make him indifferent between $10 immediately and $X in one year.

How do results from these two methods compare? Ahlbrecht & Weber (1997) and Read & Roelofsma (2003) tested both methods in within-subjects designs and found that matching yielded lower discount rates than choice. Why does this happen? One hypothesis is that in choice people are motivated to take the earlier amount, and pay relatively more attention to the delay, while for matching they are more likely to focus on the amounts and attempt to balance the two attributes (Tversky, Sattath, & Slovic, 1988). However, neither of the previously mentioned studies manipulated task order, so it is difficult to know whether experience with matching questions influenced participants' answers on the choice task. Frederick (2003) compared seven different elicitation methods for current versus future lives, in a mixed design. He also found that matching produced lower discount rates than choice, but again, order effects were not explored. He speculated that the choice task creates demand characteristics: offering the choice between different amounts of immediate and future lives implies that one ought to discount them to some extent—"otherwise, why would the experimenter be asking the question" (Frederick, 2003, p. 42). In contrast, the matching method makes no suggestions as to which amounts are appropriate.

Further evidence that choice options can bias discount rates comes from a pair of studies comparing two different variations on a choice-based measure. One version presented choices with amounts in ascending order, and the other presented amounts in descending order. The

order affected discount rates, such that participants were more patient (exhibited lower discount rates) when answering the questions in descending order (Robles & Vargas, 2008; Robles, Vargas, & Bejarano, 2009).

While these studies describe some interesting differences between the methods, they offer little guidance regarding which technique researchers ought to use to measure discount rates. One perspective would argue that because preferences are constructed, the results from different measures are equally valid expressions of people's preferences, and it is therefore impossible to choose a best measure. However, if researchers are interested in predicting and explaining real-world behaviors in other contexts, this provides an objective metric by which to make a judgment. While several studies have shown this real-world link for choice-based techniques, we are not aware of any studies examining how well matching predicts consequential decisions.

Another question concerns how these different elicitation techniques perform with non-traditional delays and outcomes. Most studies have focused on financial delays in the range of 6 months to a few years, but many consequential real-world intertemporal choices, such as retirement savings, smoking, or environmental decisions, involve much longer delays and diverse consequences.

The Current Research

We compared matching with choice-based methods of eliciting discount rates, in a mixed design. Half the participants completed matching, then choice, while the other half did the opposite order. This allowed us to analyze the data as both within and between subjects. Within each measurement technique, delays varied from one year to fifty years. Also, all participants completed both a hypothetical financial discounting scenario as well as an air quality discounting

scenario. Outcome sign was manipulated between subjects, such that half the participants considered current versus future gains, while the other half considered current versus future losses.

Within the choice-based condition, we compared two different techniques: fixed-sequence *titration* and a dynamic *multiple staircase* method. While the titration method presented participants with a pre-set list of choices, the multiple staircase method (described in detail in Appendix C) dynamically selected choice options based on participants' previous answers, to funnel in on their indifference points more efficiently. Within the titration condition, we manipulated presentation order of the future outcomes as either low-to-high or high-to-low.

At the end of the survey, we presented participants with a consequential choice between $100 today or $200 next year, and randomly paid out two participants for real money. We also asked participants whether they smoked or not, to get data on a consequential life choice.

We compared the measurement methods in four different categories: ability to detect inattentive participants, differences in central tendency and variability, ease of use for participants, and ability to predict consequential intertemporal choices. We predicted that the multiple staircase method would be best at detecting inattentive participants, because we designed it partly with this purpose in mind. We predicted that the choice-based methods would show higher discount rates than the matching method, based on the results from previous research discussed above. We predicted that the choice-based methods would be easier for participants to understand and use, based on anecdotal evidence from our own previous research indicating that participants often have a hard time with the concept of indifference, and have a hard time picking a number "out of the air" without any reference. Finally, we predicted that the

choice-based methods would be better at predicting the consequential choices, because there is a natural congruence in using choice to predict choice, and because previous studies have established the efficacy of choice-based methods but none have shown support of matching as a consequential-choice predictor.

Methods

516 participants (68% female, mean age=38, *SD*=13) were recruited online for a study on decision making and randomly assigned to an experimental condition. Participants in the gain condition read the following hypothetical scenario:

> Imagine the city you live in has a budget surplus that it is planning to pay out as rebates of $300 for **each** citizen. The city is also considering investing the surplus in endowment funds that will mature at **different possible times** in the future. The funds would allow the city to offer rebates of a different amount, to be paid at different possible times in the future. For the purposes of answering these questions, please assume that you will not move away from your current city, even if that is unlikely to be true in reality.

The full text all the scenarios can be found in Appendix A. After reading the scenario, participants then indicated their intertemporal preferences in one of three different ways. In the matching condition, participants filled in a blank with an amount that would make them indifferent between $300 immediately and another amount in the future (see Appendix B for examples of the questions using each measurement method). Participants answered questions about three different delays: 1 year, 10 years, and 50 years. Although many participants might expect to be dead in 50 years, the scenario described future gains that would benefit everyone in their city, so it was hoped that those future gains would still have meaning to participants. In the titration condition, participants made a series of choices between immediate and future amounts, at each delay. Because the same set of choice options was presented for each delay (see

Appendix B for the list of options), the choice set offered a wide range of values, to allow for high discount rates at long delays. The order of the future amounts was balanced between participants, such that half answered lists with amounts going from low to high (as in Appendix B), and others were presented with amounts going from high to low.

In the multiple staircase, participants also made a series of choices between immediate and future amounts. Unlike the simple titration method, these amounts were selected dynamically, funneling in on the participant's indifference point. Choices were presented one at a time (unlike titration, which presented all choices on one page). Also unlike titration, the questions from the three delays were interleaved in a random sequence. The complete multiple staircase method is described in detail in Appendix C.

In all conditions progress in the intertemporal choice task was indicated with a progress bar. Also in all conditions, participants could refer back to the scenario as they answered the questions. After completing the intertemporal choice task, participants were asked "What things did you think about as you answered the previous questions? Please give a **brief** summary of your thoughts." This allowed us to collect a qualitative measure of the process participants went through while responding to the questions.

Next, participants answered the same intertemporal choice scenario using a different measurement method. Those who initially were given a choice-based measure (titration or multiple staircase) subsequently completed a matching measure, while those who began with matching then completed a choice-based measure. In other words, all participants completed a matching measure, either before or after completing a choice-based measure. This design was

chosen because the matching measure is relatively quick for participants to complete, so it was easy to include it in all conditions.

Subsequently, participants were given an attention check, very similar to the Instructional Manipulation Check (IMC; Oppenheimer, Meyvis, & Davidenko, 2009), which ascertained whether participants were reading instructions.

After that, participants read an environmental discounting scenario, the full text of which can be found in Appendix A. This asked them to choose between an immediate amount of money, and a change in air quality that would happen at different times. In other words, participants were asked how much the change in air quality was worth, depending on the delay. There were four different delays in the air quality scenario: immediate, 1 year, 10 years, and 50 years. Again, it was hoped that the 50-year delayed change in air quality would still be meaningful to participants because it would affect the residents of the city they lived in. Participants were instructed to answer based on an assumption that they would remain living in their current city.

Next, participants completed demographics, including a question about whether they smoke. Finally, participants completed a consequential measure of intertemporal choice, in which they chose between receiving $100 immediately or $200 in one year (note that participants in the loss condition still chose between two gains in this case, due to the fact that it would be difficult to execute losses for real money). Participants were informed that two people would be randomly selected and have their choices paid out for real money, and this indeed happened.

Results and Discussion

Detecting inattentive participants

In most psychology research, and especially in online research, there are often participants who do not pay much attention or do not respond carefully. It is helpful, therefore, if measurement methods can detect these participants. The multiple-staircase method had two built-in check questions (described above) to detect such participants. The titration method can also detect inattention, by looking for instances of switching back and forth, or switching perversely. For example, if a participant preferred $475 in one year over $300 today, but preferred $300 today over $900 in one year, this would be a sign of inattention. Titration cannot, however, differentiate between "good" participants and those who learn what a "good" pattern of choices looks like and reproduce such a pattern for later questions, without carefully considering each subsequent question individually. It is nearly impossible for a single matching measure to detect inattention, but with multiple measures presented at different time points, matching may identify those participants who show a non-monotonic effect of time. For example, if the one-year indifference point (with respect to $300 immediately) is $5000, the ten year indifference point is $600, and the fifty year indifference point is $500000, this might be evidence of inattention.

As described above, each participant also completed another attention check, which was very similar to the Instructional Manipulation Check (Oppenheimer, et al., 2009). As this measure has been empirically shown to be effective for detecting inattentive participants, we compared the ability of each measurement method to predict IMC status.

Correlations between the IMC and each measure's test of attention revealed that while neither matching, r=.07, p>.1, nor titration, r=.06, p>.1 were able to detect inattentive participants, the multiple staircase method had modest success, r=.19, p<.05. Overall, then, no method was particularly effective at detecting inattentive participants, but the multiple-staircase method was superior to the others.

For all of the following analyses, we only wanted to compare those participants who were paying attention and reading instructions. Therefore, we excluded those participants who failed the IMC, leaving 316 participants for further analysis.

Differences in central tendency and variability

Indifference points were computed for each participant and time delay as follows: in matching, the number given by participants was used directly. In titration, the average of the values around the switch point was used. For example, if a participant preferred $300 immediately over $475 in ten years, but preferred $900 in ten years over $300 immediately, the participant was judged to be indifferent between $300 immediately and $687.50 in ten years. In multiple staircase, the average of the established upper bound and lower bound was used, in a similar manner to titration. These indifference points were then converted to discount rates, using three different popular equations: exponential, hyperbolic, and area under the curve. The continuously compounded exponential discount rate (Samuelson, 1937) is calculated as $V=Ae^{-kD}$, where V is the present value, A is the future amount, *e* is the constant (2.718...), D is the delay in years, and k is the discount rate. This is the normative model of discounting. The hyperbolic model (Mazur, 1987) is calculated by $V=A / (1+kD)$, where V is the present value, A is the future amount, D is the delay (often in years), and k is the discount rate. This hyperbolic model has been found to descriptively model discounting data better than the exponential model. The third discounting equation we used was the area under the curve (AUC) metric, which sets the present value and the longest delay each equal to 1 and computes the fraction that future amounts are worth by computing the area under the curve of the indifference points (Myerson, Green, & Warusawitharana, 2001), by summing $(x_2-x_1)*[(y_2+y_1)/2]$ for each indifference point, where x_1 is the sooner time point, x_2 is the later time point, y_1 is the sooner amount, and y_2 is the later

amount. This generally yields a fraction between 1 and 0 where lower numbers indicate more discounting. The AUC is supposedly a theory-free measure of discounting behavior. It is important to note that in addition to the differences in theory and explanatory power, these three models also differ in the ways that they transform the data. For example, while the AUC minimizes extreme discounting, it *magnifies* extreme negative discounting. For example, if someone reports that they would pay $5 for an increase in air quality starting today, but $15 for an increase in air quality starting in one year, this would yield an AUC of 10, which is quite extreme given that the scale normally yields values between 1 and 0.

Because order effects were observed (which we describe below), the majority of the analyses to follow will focus on the first measurement method that participants completed. This leaves n=154 in the matching condition, n=82 in the titration condition, and n=80 in the multiple staircase condition. Discount rates from each condition are summarized in Table 1. Because skew and outliers were sometimes pronounced, this table lists median and interquartile range in addition to mean and standard deviation.

As is clear from Table 1, many different stories can be told from the data depending on the measurement method and discounting model used. For example, when using the exponential formula, the titration method shows the largest standard deviation for financial gains, whereas when using the AUC formula, the matching method shows the largest standard deviation. Yet, there are some consistencies across measurements methods and equations. In *all* cases, whether using matching, titration or multiple staircase; exponential discounting, hyperbolic discounting or AUC; means or medians; financial gains were always discounted more than financial losses (when looking at Table 1, recall that AUC is reverse scored relative to the other measures).

Table 1

Means, standard deviations, medians, and interquartile ranges (IQRs) for three methods of measuring discount rates (matching, multiple staircase, and titration) for financial gains and losses. Discount rates are summarized for three popular discounting equations: the continuously compounded exponential model, the hyperbolic model, and the area under the curve (AUC). Note that for the first two models, higher numbers mean more discounting, while with the AUC lower numbers mean more discounting.

Financial Outcome	Exponential Discount Rate				Hyperbolic Discount Rate				Area Under the Curve			
	Mean	SD	Median	IQR	Mean	SD	Median	IQR	Mean	SD	Median	IQR
matching, gain	.24	.21	.23	.23	2.47	8.53	0.74	1.11	0.43	1.38	0.17	0.20
m-stairs, gain	.46	.38	.34	.37	6.03	16.01	2.55	2.14	0.13	0.08	0.11	0.12
titration, gain	.41	.52	.18	.30	10.27	34.43	1.59	3.07	0.16	0.10	0.15	0.15
matching, loss	.08	.33	.08	.21	0.53	1.47	0.15	.41	4.93	32.77	0.44	0.36
m-stairs, loss	.22	.26	.12	.23	1.29	2.92	0.29	.53	0.35	0.27	0.25	0.32
titration, loss	.27	.51	.10	.24	3.69	14.29	0.14	.58	0.60	0.65	0.40	0.61

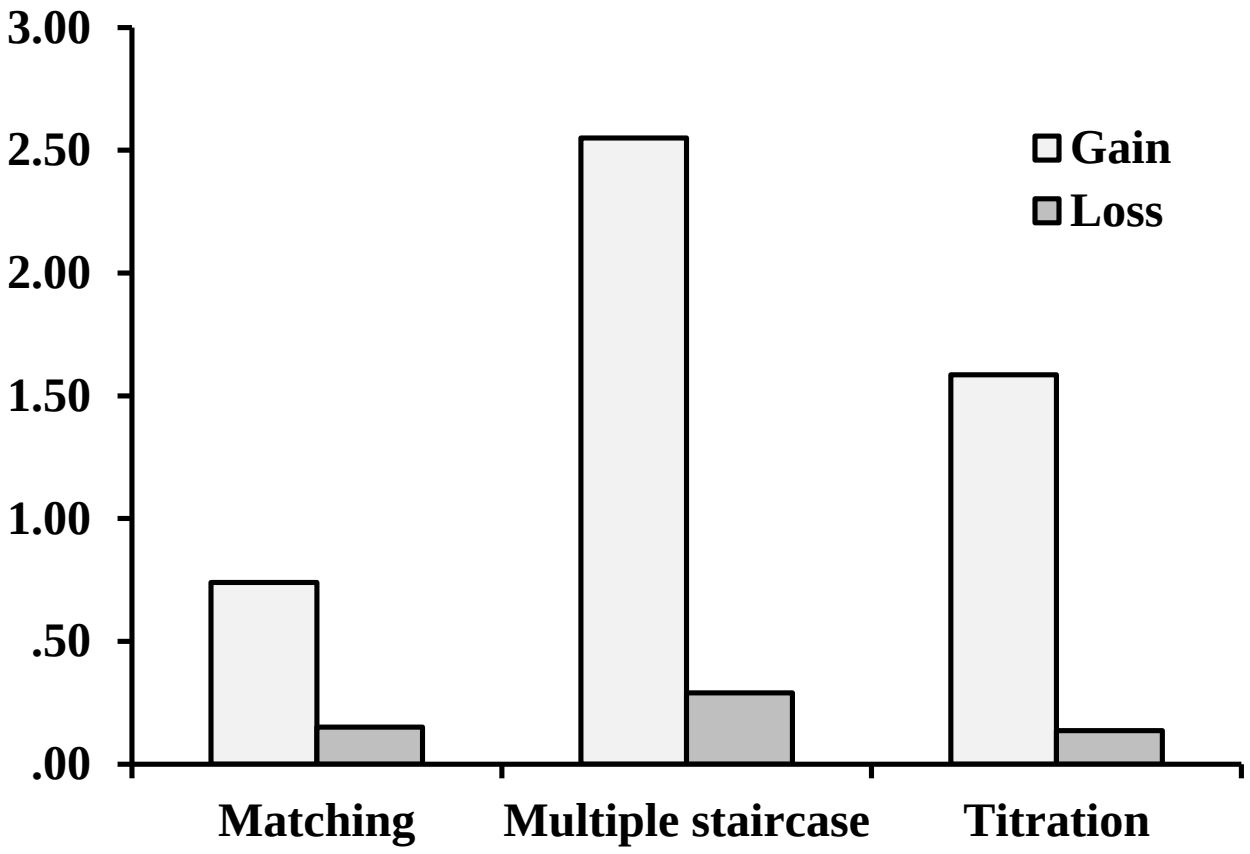
Because most psychological studies of discounting employ the hyperbolic model, our further analyses will focus on this model. As this model is subject to skew and outliers, we will primarily use non-parametric statistics. We feel that this approach (median hyperbolic discount rates) best represents the overall pattern of data.

As seen in Figure 1, discount rates measured with the choice-based methods (titration and multiple staircase) were generally higher than discount rates measured with matching, and this was particularly true for gains. A Kruskal-Wallis non-parametric ANOVA run on the gain data confirmed significant effect of measurement method on discount rates, $p<.001$, and another

Kruskal-Wallis test run on losses found a significant effect of measurement method there as well, $p<.05$.

Figure 1

Median hyperbolic discount rates for financial gains and losses as measured with matching, multiple staircase, or titration.

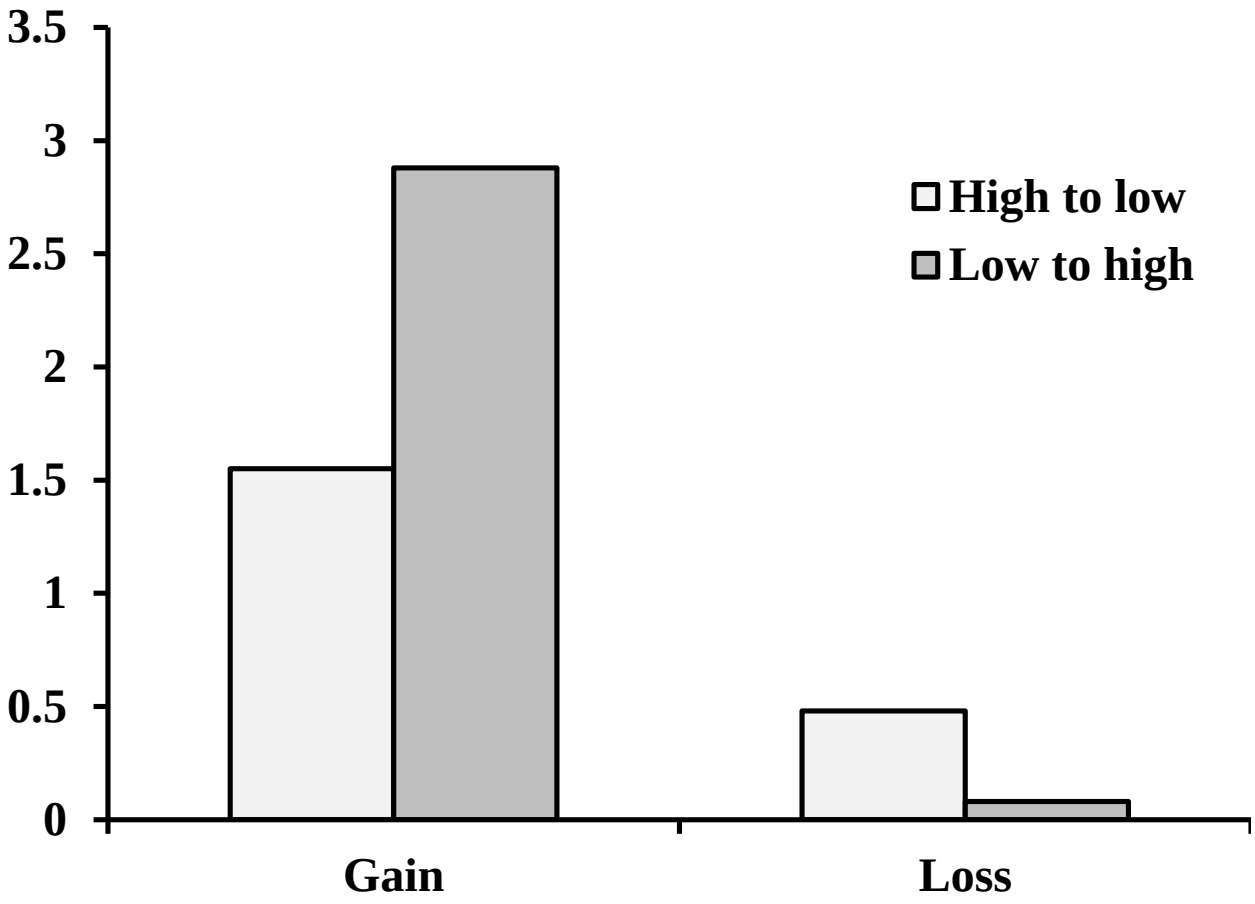


We hypothesized that these differences in discount rates were partly a function of anchoring or demand characteristics. In other words, the extreme options sometimes presented to participants (such as a choice between $300 today and $85,000 in one year) may have suggested that these were reasonable choices, and so encouraged higher discount rates. For comparison, an earlier study from our same lab (Hardisty & Weber, 2009, Study 1), drawing on the same

participant pool and using titration for financial outcomes, presented participants with a much smaller range of options ($250 today vs. $230 to $410 in one year) and yielded much lower discount rates: 0.28 for gains, and 0.04 for losses, compared with 1.59 for gains and 0.14 for losses in the present study. It seems, then, the range of options presented to participants affected their discount rates by suggesting reasonable options as well as by restricting what participants could or could not actually express. We also tested for the influence of the options presented to participants by comparing the two orderings, high-to-low and low-to-high. As summarized in Figure 2, this ordering manipulation did indeed affect responses. In particular, participants tended to perseverate on the side they started out on. So, for example, a participant who began with a choice between receiving $300 immediately or $250 in the future would begin by choosing the immediately amount, and stay with it for a little while, thus exhibiting a high discount rate (i.e., greater impatience). In contrast, a participant who was first presented with receiving $300 immediately versus $85,000 in the future would begin by selecting the future option, and continue to choose more future options, thus showing a low discount rate. Losses showed the same effect, which translates into discount rates in the opposite manner. For example, a participant who begins with a choice between paying $300 now or $250 in the future will likely choose the future amount, and continue choosing to pay in the future, which ultimately leads to a higher discount rate.

Figure 2

Median hyperbolic discount rates for financial gains and losses as measured with titration, broken down by the order in which future amounts were presented: high to low, or low to high.

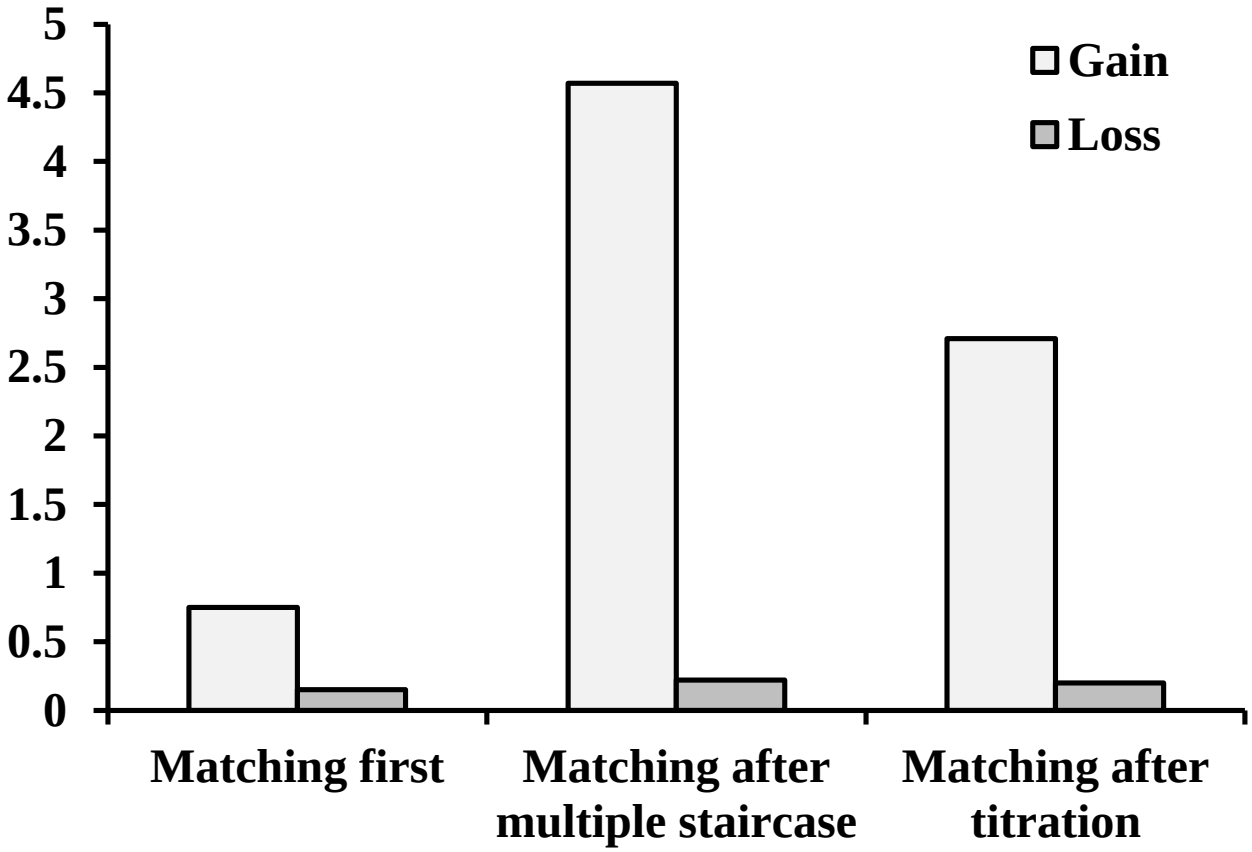


A Mann-Whitney U test comparing the high-to-low and low-to-high orderings for gains was significant, $p<.01$. A similar test comparing the two orderings for losses was not significant, $p>.1$, but was in the predicted direction.

While discount rates were generally much higher when using the choice-based methods, we believe that this was due to the large range of options that we presented to participants, and it would be possible to obtain the opposite pattern of results if only small amounts were used. Further evidence for this comes from a within-subject analysis comparing the different methods: although all participants completed a matching measure, some did it before a choice method, and some did it after a choice method. Comparing these participants reveals a significant effect of order, as seen in Figure 3.

Figure 3

Median hyperbolic discount rates for financial gains and losses as measured with matching, broken down by whether participants did the matching task before or after one of the choice-based measurement methods.



A Kruskal-Wallis test confirmed that participants who answered the choice-based measures first gave higher discount rates in response to the following matching questions than those who did matching first, both for gains, $p<.001$, and for losses, $p=.05$.

Just as the mean and median discount rates yielded by the choice-based methods were higher than those from the matching method, so too was the spread of the distributions from the choice-based methods larger. The interquartile range (IQR) from the matching method for gains was 1.1, compared with 2.1 from multiple staircase and 3.1 from titration. Similarly, the IQR for matching losses was only .41, compared with .53 from multiple staircase and .58 from titration.

It is likely that the same factors that led to the higher medians in the choice-based methods also produced the greater IQR.

Discount rates for the air quality outcomes were computed by comparing how much participants valued the immediate change in air quality versus the future change in air quality. For example, if someone would pay $100 to avoid an immediate deterioration in air quality, but only pay $90 to avoid a deterioration starting in one year, then the hyperbolic discount rate for that participant would be (100-90)/(90*1), or .11. Perhaps as a result of the immediate amount being variable in addition to the future amounts, this made the variances in discount rates quite large for the air quality outcomes, as seen in Table 2. Also, some participants were confused, particularly in the matching condition (discussed below), so the discount rates in Table 2 probably do not represent participants' true preferences. Therefore, we will not analyze them very closely, except to note that the same general trends are seen here as well as for the financial outcomes: choice-based methods generally produced higher discount rates than matching, and often had higher variability as well.[1]

[1] The astute reader may notice that overall, discount rates for the air quality outcomes are lower than those for the financial outcomes. This may seem to contradict Hardisty & Weber (2009), who found little difference in discount rates between financial and environmental outcomes. We believe this is due to the fact the 2009 paper used within-domain measures of discounting (air quality now vs. a different amount of air quality in the future), while the present study uses a between-domain measure of discounting (how many dollars is air quality worth now vs. later). An unpublished study by our lab showed that this is indeed the case. We speculate that many participants hold a moral conviction that environmental things ought to be worth the same amount no matter when they happen, and so many participants show discount rates of zero when using this procedure. In contrast, the 2009 paper showed that the vast majority of participants prefer a smaller improvement in air quality now over a larger one in the future, thus showing standard temporal discounting behavior.

Table 2

Means, standard deviations, medians, and interquartile ranges (IQRs) for three methods of measuring discount rates (matching, multiple staircase, and titration) for air quality gains and losses. Discount rates are summarized for three popular discounting equations: the continuously compounded exponential model, the hyperbolic model, and the area under the curve (AUC). Note that for the first two models, higher numbers mean more discounting, while with the AUC lower numbers mean more discounting.

Air Quality Outcome	Continuously Compounded Exponential Discount Rate				Hyperbolic Discount Rate				Area Under the Curve			
	Mean	SD	Median	IQR	Mean	SD	Median	IQR	Mean	SD	Median	IQR
matching, gain	-.13	.40	.00	.33	0.05	0.36	0.00	.29	17.24	47.30	1.00	8.53
m-stairs, gain	.35	.50	.21	.49	68.53	319.7	1.16	15.63	0.30	0.34	0.14	.38
titration, gain	.18	.27	.05	.35	1.35	3.62	0.31	.69	0.51	0.36	0.41	.39
matching, loss	-.02	.25	.00	.20	0.08	0.30	0.00	.21	8.56	50.35	1.00	1.87
m-stairs, loss	.11	.29	.02	.08	3.53	15.22	0.03	.19	0.73	0.56	0.72	.67
titration, loss	.04	.22	.00	.05	9.12	58.53	0.00	.15	1.51	3.18	1.00	.53

Overall, then, we have evidence that the options presented to participants in the choice-based methods affected the discount rates that they expressed. While matching has the advantage of not providing any anchors or suggestions to participants, it is nonetheless still quite susceptible to influence from other sources. This is not a particularly novel finding, as theories and findings of constructed preference (Johnson, Haubl, & Keinan, 2007; Weber et al., 2007) and coherent arbitrariness (Ariely, Loewenstein, & Prelec, 2003) are plentiful. However, it has not received much attention in intertemporal choice. Many differences in discount rates between

studies may be explained by differences in the amount and order of options that experimenters presented to participants.

Ease of use for participants

Another factor in determining which method to use is how easy it is for participants to complete. Qualitative evidence from piloting and from participants' comments indicated that participants often had a hard time understanding and answering the matching questions. They had a hard time picking a number "out of the air," and also had a hard time understanding the concept of indifference points. While the titration method was relatively easy for participants to answer, they found the multiple staircase method quite onerous. Even though the number of questions was smaller than for titration, participants perceived it to be longer. It was difficult for participants to switch back and forth between different timescales, and participants had trouble answering the later questions, which were often quite near their indifference points and so were difficult to decide on.

Quantitative evidence to support these observations came from participants' responses to the environmental scenario. The rational response is to value future improvements or deterioration in air quality less than immediate improvements or deterioration in air quality. For example, if someone is willing to pay $200 for an immediate improvement in air quality, then they should not be willing to pay more than $200 for an improvement that would begin in fifty years. We calculated the proportion of participants that valued the fifty-year change more than the immediate change, and found it to be 38% with matching, 11% with multiple staircase, and 7% with titration. Pairwise proportion tests indicated that both of the choice-based methods showed less confused responses than the matching method, $p<.001$. Qualitative data reinforced the assumption that the matching participants were indeed making confused mistakes when

responding to the environmental scenario, rather than expressing their true preferences. For example, one typical participant wrote "This was confusing to me. I would pretty much always take better air quality over a financial incentive. I wasn't clear whether this would mean putting a high value on the rebate option now or in the future." Our interpretation is that it is difficult for participants to pull dollar values out of the air, and while this is somewhat manageable when participants only have to think about different amounts of money at different times, it becomes extremely difficult to do when participants have to consider tradeoffs between air quality and money at different points in time. Therefore, although 38% of the participants in our matching sample showed negative discount rates, we believe that these were nearly all errors in responding.

Predicting consequential choices

Perhaps most importantly, researchers are interested in understanding real, consequential choices that people make. We therefore compared the ability of discount rates from the hypothetical scenarios to predict two consequential choices. First, we used the 1-year hyperbolic discount rate to predict whether participants would choose to receive $100 today or $200 in the future. Due to skew and outliers in the distributions, we used the non-parametric Spearman's rho. As seen in Table 2, the correlations were always positive, meaning that participants with higher discount rates were more likely to choose the immediate $100. The choice-based methods clearly outperformed the matching method, and this makes intuitive sense: predicting a choice will be easier with a choice-based measure than a fill-in-the-blank measure. The better predictive power may also stem from the fact that participants found it easier to understand and respond to the choice-based measures of discounting.

We then looked at the ability of these discount rates to predict a real-life choice: whether each participant was a smoker or not. As seen in Table 2, the choice-based methods were sometimes able to predict this (with higher discount rates correlated with smoking), while the matching method was not. Oddly, the multiple-staircase method was better at predicting smoking rates with discount rates for gains, while titration was better at predicting using discount rates for losses. We don't have a good explanation for this difference.

Table 3

*Non-parametric correlations (Spearman's rho) between hypothetical discount rates measured in different ways and consequential intertemporal choices. The † symbol indicates p<.1 two-tailed, * indicates p<.05, and ** indicates p<.01.*

	Choosing a $100 gain now over $200 in one year	Smoking
matching, gain	.07	-.04
m-stairs, gain	.33*	.40**
titration, gain	.67**	.00
matching, loss	.16	.16
m-stairs, loss	.28†	.06
titration, loss	.26†	.41**

These results suggest that researchers who are interested in predicting consequential intertemporal choices should employ choice-based methods. It's possible that matching might be better for predicting those real-life situations that require the person to name a number, such as deciding how much to contribute to a retirement plan, or how quickly to pay down a debt. This remains to be tested in future research.

Conclusions

Choice-based measures of discounting are a double-edged sword, to be used carefully. On one hand, they outperform matching at predicting consequential intertemporal choices. On the other hand, the options (and order of options) that researchers use will influence participants' answers, so experimental design and interpretation must be done with care. Differences in discount rates observed between studies may be attributed to differences in elicitation technique, consistent with long-established research on risky choice that has come to the same conclusion (Lichtenstein & Slovic, 1971; Tversky, et al., 1988).

Overall, discount rates elicited from choice-based methods were higher than those from matching, consistent with previous research (Ahlbrecht & Weber, 1997; Frederick, 2003; Read & Roelofsma, 2003). We agree with Frederick (2003) that this probably happens because the choice-based methods implicitly suggest high discount rates to participants, while matching has less demand characteristics. However, it should easily be possible to design a study with a choice-based measure of discounting that would elicit lower discount rates than matching techniques, simply by choosing a range of choice options that is extremely low. When doing within-subjects analysis, we found strong order effects; participants gave very different responses to the matching questions depending on whether they completed them before or after a choice-based method. Therefore, future research on methods should be careful to account for this.

In in comparison with the standard, fixed-sequence titration method, we did not find compelling advantages for the multiple-staircase method we developed. This is consistent with

another recent study on dynamic versus fixed sequence choice, which also found no major differences (Rodzon, Berry, & Odum, 2011). In some ways, it is disappointing that our attempt to improve measurement was unsuccessful. However, the good news is that the simple titration measure, which is much more convenient to implement, still remains one of the best methods.

While we have focused on choice and matching techniques, because have been most commonly used in the literature, it should be noted that many other techniques have recently been tested and compared, including intertemporal allocation, evaluations of sequences, and intertemporal auctions (Frederick & Loewenstein, 2008; Guyse & Simon, working paper; Olivola & Wang, working paper). All of these investigations have found difference in discount rates based on the elicitation technique.

In terms of best practices for studying temporal discounting, our recommendation for now is to use the measurement method most congruent with the real-world phenomenon one hopes to explain or understand. When in doubt, a choice-based method should be used, because it has shown the best predictive power so far. However, care must always be taken when considering the choice options presented to participants, as these will undoubtedly influence the observed discount rates.

References

Ahlbrecht, M., & Weber, M. (1997). An empirical study on intertemporal decision making under risk. *Management Science, 43*(6), 813-826. doi: 10.1287/mnsc.43.6.813

Ariely, D., Loewenstein, G., & Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. *The Quarterly Journal of Economics, 118*, 73-105. doi: 10.1162/00335530360535153

Chabris, C. F., Laibson, D., Morris, C. L., Schuldt, J. P., & Taubinsky, D. (2008). Individual laboratory-measured discount rates predict field behavior. *Journal of Risk and Uncertainty, 37*. doi: 10.1007/s11166-008-9053-x

Frederick, S. (2003). Measuring intergenerational time preference: Are future lives valued less? *Journal of Risk & Uncertainty, 26*, 39-53. doi: 10.1023/A:1022298223127

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

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

Guyse, J. L., & Simon, J. (working paper). Consistency among elicitation techniques for intertemporal choice: A within-subjects investigation of the anomalies.

Hardisty, D. J., & Weber, E. U. (2009). Discounting future green: Money versus the environment. *Journal of Experimental Psychology: General, 138*(3), 329-340. doi: 10.1037/a0016433

Johnson, E. J., Haubl, G., & Keinan, A. (2007). Aspects of endowment: A query theory of value. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 461-474. doi: 10.1037/0278-7393.33.3.461

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, *89*(1), 46-55. doi: 10.1037/h0031207

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazure, J. A. Nevin & H. Rachlin (Eds.), *Quantitative analyses of*

behavior: Vol. 5. The effect of delay and intervening events on reinforcement value (pp. 55-73). Hillsdale, NJ: Erlbaum.

Meier, S., & Sprenger, C. (2010). Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics, 2*(1), 193-210. doi: 10.1257/app.2.1.193

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

Olivola, C. Y., & Wang, S. W. (working paper). Patience auctions: The impact of time vs money bidding on elicited discount rates.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*(4), 867-872. doi: 10.1016/j.jesp.2009.03.009

Read, D., & Roelofsma, P. (2003). Subadditive versus hyperbolic discounting: A comparison of choice and matching. *Organizational Behavior & Human Decision Processes, 91*, 140-153. doi: 10.1016/S0749-5978(03)00060-8

Reimers, S., Maylor, E. A., Stewart, N., & Chater, N. (2009). Associations between a one-shot delay discounting measure and age, income education and real-world impulsive behavior. *Personality and Individual Differences, 47*, 973-978. doi: 10.1016/j.paid.2009.07.026

Robles, E., & Vargas, P. A. (2008). Parameters of delay discounting assessment: Number of trials, effort, and sequential effects. *Behavioral Processes, 78*(285-290). doi: 10.1016/j.beproc.2007.10.012

Robles, E., Vargas, P. A., & Bejarano, R. (2009). Within-subject differences in degree of delay discounting as a function of order of presentation of hypothetical cash rewards. *Behavioral Processes, 81*, 260-263. doi: 10.1016/j.beproc.2009.02.018

Rodzon, K., Berry, M. S., & Odum, A. L. (2011). Within-subject comparison of degree of delay discounting using titrating and fixed sequence procedures. *Behavioural Processes, 86*, 164-167. doi: 10.1016/j.beproc.2010.09.007

Samuelson, P. (1937). A note on measurement of utility. *Review of Economic Studies, 4*, 155-161. doi: 10.2307/2967612

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95*(3). doi: 10.1037/0033-295X.95.3.371

Weber, E. U., Johnson, E. J., Milch, K. F., Chang, H., Brodscholl, J. C., & Goldstein, D. G. (2007). Asymmetric discounting in intertemporal choice. *Psychological Science, 18*(6), 516-523. doi: 10.1111/j.1467-9280.2007.01932.x

Appendix A: Full text of the discounting scenarios

Financial Gain

Imagine the city you live in has a budget surplus that it is planning to pay out as rebates of $300 for **each** citizen. The city is also considering investing the surplus in endowment funds that will mature at **different possible times** in the future. The funds would allow the city to offer rebates of a different amount, to be paid at different possible times in the future. For the purposes of answering these questions, please assume that you will not move away from your current city, even if that is unlikely to be true in reality.

Financial Loss

Imagine the city you live in has a budget shortfall that it is planning to cover through a one-time fee of $300 for **each** citizen. The city is also considering covering the shortfall using fixed-interest bonds that will mature at **different possible times** in the future. Offering these bonds would require the city to charge the citizens a different amount, to be paid when the bonds mature. For the purposes of answering these questions, please assume that you will not move away from your current city, even if that is unlikely to be true in reality.

Environmental Gain

Imagine the current air quality (measured by number and size of particulates) in your area is neither particularly good nor especially bad. The local government has a budget surplus that it will either return to the citizens as rebates, or spend to enact various policy and infrastructure changes that will lead to a permanent improvement in air quality. Once the changes are put into place, the air will feel surprisingly clean and fresh.

Policy changes will include stricter emissions standards for factories and power plants; the city will compensate those factories and power plants for any costs incurred. Infrastructure changes will include using a fleet of cleaner-burning, more fuel-efficient vehicles in place of those currently used in the public transportation system and by city employees.

We are **not** interested in how you feel about specific measures meant to improve air quality. Rather, we are interested in knowing how much this improved air quality would be worth to you, depending on when the change is implemented. The following questions will ask about your

preference between receiving a sum of money as a rebate now, or having noticeably improved air quality starting at different possible times, now or in the future. For the purposes of these questions, please assume you will continue to live in your current city, even if that is unlikely to be true in reality.

Environmental Loss

Imagine the current air quality (measured by number and size of particulates) in your area is neither particularly good nor especially bad. The local government has a budget shortfall that it will either cover by charging the citizens, or reduce spending on various policy and infrastructure repairs, leading to a permanent deterioration in air quality. Once the changes are put into place, the air will feel surprisingly dirty and stale.

Policy changes will include weaker emissions standards for factories and power plants; the city will earn more in taxes from those factories and power plants, as their profits will increase under the weaker standards. Infrastructure changes will include using a fleet of cheap, less fuel-efficient vehicles in place of those currently used in the public transportation system and by city employees.

We are **not** interested in how you feel about specific measures affecting the air quality. Rather, we are interested in knowing how much it would be worth to you to avoid this worsened air quality, depending on when the change is implemented. The following questions will ask about your preference between paying a sum of money as a one-time fee now, or having noticeably worsened air quality starting at different possible times, now or in the future. For the purposes of

these questions, please assume you will continue to live in your current city, even if that is unlikely to be true in reality.

Appendix B: Sample questions from each of the measurement methods and scenario types

Matching, Financial Gain

How much would a rebate **one year** from now have to be in order to make it equally attractive as $300 **now**?

Please fill in the amount that would make the following options equally attractive.

A. Receive $300 **immediately**.

B. Receive $____ **one year** from now.

Matching, Financial Loss

How much would a tax **one year** from now have to be in order to make it as unattractive as paying $300 **now**?

Please fill in the amount that would make the following options equally unattractive.

A. Pay $300 **immediately**.

B. Pay $____ **one year** from now.

Titration, Financial Gain

Please choose the option that you prefer in each pair.

What if the rebate were to be paid **one year** from now?

A1.	Receive $300 **immediately**	Receive $250 **one year** from now
A2.	Receive $300 **immediately**	Receive $475 **one year** from now
A3.	Receive $300 **immediately**	Receive $900 **one year** from now
A4.	Receive $300 **immediately**	Receive $1,750 **one year** from now
A5.	Receive $300 **immediately**	Receive $3,300 **one year** from now
A6.	Receive $300 **immediately**	Receive $6,400 **one year** from now
A7.	Receive $300 **immediately**	Receive $12,000 **one year** from now
A8.	Receive $300 **immediately**	Receive $23,500 **one year** from now
A9.	Receive $300 **immediately**	Receive $45,000 **one year** from now
A10.	Receive $300 **immediately**	Receive $85,000 **one year** from now

Titration, Financial Loss

Please choose the option that you prefer in each pair.

What if the tax were to be paid **one year** from now?

A1.	Pay $300 **immediately**	Pay $250 **one year** from now
A2.	Pay $300 **immediately**	Pay $475 **one year** from now
A3.	Pay $300 **immediately**	Pay $900 **one year** from now
A4.	Pay $300 **immediately**	Pay $1,750 **one year** from now
A5.	Pay $300 **immediately**	Pay $3,300 **one year** from now
A6.	Pay $300 **immediately**	Pay $6,400 **one year** from now
A7.	Pay $300 **immediately**	Pay $12,000 **one year** from now
A8.	Pay $300 **immediately**	Pay $23,500 **one year** from now
A9.	Pay $300 **immediately**	Pay $45,000 **one year** from now
A10.	Pay $300 **immediately**	Pay $85,000 **one year** from now

Multiple Staircase, Financial Gain

Which option do you prefer:

Receive $300 **now** OR Receive $7,700 **one**

	year from now

Multiple Staircase, Financial Loss

Which option do you prefer:

Pay $300 **now**	OR	Pay $7,736 **one year from now**

Matching, Environmental Gain

What amount of money now would be as valuable to you as getting improved air quality starting **one year from now**? In other words, how much would a rebate have to be in order to make it difficult or impossible for you to choose whether you would prefer getting that amount of money immediately or getting the cleaner air in one year?

Please fill in the amount that would make the following options equally attractive.

(NOTE: $0 would indicate that improved air quality is worthless to you.)

A. Improved air quality starting **one year from now**.

B. Receive $____ **immediately**.

Matching, Environmental Loss

Paying what amount of money now would be as costly to you as suffering worse air quality starting **one year from now**? In other words, how much would a tax have to be in order to make it difficult or impossible for you to choose whether you would prefer paying that amount of money immediately or suffering the dirty air in one year?

Please fill in the amount that would make the following options equally unattractive.

(NOTE: $0 would indicate that air quality is worthless to you.)

A. Worse air quality starting **one year from now**.

B. Pay $____ **immediately**.

Titration, Environmental Gain

What if the improved air quality were to start **one year from now**?

B1.	Receive $20 **immediately**.	Permanently improved air quality starting **one year from now**.
B2.	Receive $50 **immediately**.	Permanently improved air quality starting **one year from now**.
B3.	Receive $130 **immediately**.	Permanently improved air quality starting **one year from now**.
B4.	Receive $325 **immediately**.	Permanently improved air quality starting **one year from now**.
B5.	Receive $800 **immediately**.	Permanently improved air quality starting **one year from now**.
B6.	Receive $2,100 **immediately**.	Permanently improved air quality starting **one year from now**.
B7.	Receive $5,200 **immediately**.	Permanently improved air quality starting **one year from now**.
B8.	Receive $13,000 **immediately**.	Permanently improved air quality

		starting **one year from now**.
B9.	Receive $33,000 **immediately**.	Permanently improved air quality starting **one year from now**.
B10.	Receive $85,000 **immediately**.	Permanently improved air quality starting **one year from now**.

Titration, Environmental Loss

What if the worse air quality were to start **one year from now**?

B1.	Pay $20 **immediately**.	Permanently worse air quality starting **one year from now**.
B2.	Pay $50 **immediately**.	Permanently worse air quality starting **one year from now**.
B3.	Pay $130 **immediately**.	Permanently worse air quality starting **one year from now**.
B4.	Pay $325 **immediately**.	Permanently worse air quality starting **one year from now**.
B5.	Pay $800 **immediately**.	Permanently worse air quality starting **one year from now**.
B6.	Pay $2,100 **immediately**.	Permanently worse air quality starting **one year from now**.

B7.	Pay $5,200 **immediately**.	Permanently worse air quality starting **one year from now**.
B8.	Pay $13,000 **immediately**.	Permanently worse air quality starting **one year from now**.
B9.	Pay $33,000 **immediately**.	Permanently worse air quality starting **one year from now**.
B10.	Pay $85,000 **immediately**.	Permanently worse air quality starting **one year from now**.

Multiple Staircase, Environmental Gain

What if the improved air quality were to start **one year from now**?

B1.	Receive $20 **immediately**.	Permanently improved air quality starting **one year from now**.
B2.	Receive $50 **immediately**.	Permanently improved air quality starting **one year from now**.
B3.	Receive $130 **immediately**.	Permanently improved air quality starting **one year from now**.
B4.	Receive $325 **immediately**.	Permanently improved air quality starting **one year from now**.
B5.	Receive $800 **immediately**.	Permanently improved air quality

		starting **one year from now**.
B6.	Receive $2,100 **immediately**.	Permanently improved air quality starting **one year from now**.
B7.	Receive $5,200 **immediately**.	Permanently improved air quality starting **one year from now**.
B8.	Receive $13,000 **immediately**.	Permanently improved air quality starting **one year from now**.
B9.	Receive $33,000 **immediately**.	Permanently improved air quality starting **one year from now**.
B10.	Receive $85,000 **immediately**.	Permanently improved air quality starting **one year from now**.

Multiple Staircase, Environmental Loss

What if the worse air quality were to start **one year from now**?

B1.	Pay $20 **immediately**.	Permanently worse air quality starting **one year from now**.
B2.	Pay $50 **immediately**.	Permanently worse air quality starting **one year from now**.
B3.	Pay $130 **immediately**.	Permanently worse air quality starting **one year from now**.

B4.	Pay $325 **immediately**.	Permanently worse air quality starting **one year from now**.
B5.	Pay $800 **immediately**.	Permanently worse air quality starting **one year from now**.
B6.	Pay $2,100 **immediately**.	Permanently worse air quality starting **one year from now**.
B7.	Pay $5,200 **immediately**.	Permanently worse air quality starting **one year from now**.
B8.	Pay $13,000 **immediately**.	Permanently worse air quality starting **one year from now**.
B9.	Pay $33,000 **immediately**.	Permanently worse air quality starting **one year from now**.
B10.	Pay $85,000 **immediately**.	Permanently worse air quality starting **one year from now**.

Appendix C: The multiple staircase method

For the financial scenario, the future amount each staircase was bounded by $250 on the low end, and $100,000 on the high end. The immediate amount was always fixed at $300. Each staircase began with a choice between $300 immediately and an amount in the future that was roughly 7.5% of the maximum amount ($100,000). This was chosen based on pretesting, determining that this would reach indifference points quickly for most participants. The actual

future amount was jittered by a random amount (up to 1% greater or less than the desired amount) and rounded to the nearest dollar so that amounts would not be exactly the same among the various staircases. Therefore, the first question most participants saw was something like a choice between $300 immediately or $7548 in 1 year.

Subsequent questions in each staircase were chosen dynamically based on the participant's response to the previous question. The future amount was chosen to be 80% of the between the previous amount and the maximum or minimum, as appropriate. For example, if, in the first question, the participant preferred $300 today over $7,548 in the future, the next choice might be between $300 today and $1,709 in 1 year (again, the future amount is jittered). Alternately, if the participant initial preferred the future $7,548 over $300 today, the next question might be a choice between $300 today and $81,510 in one year. This 80% method was chosen rather than bisection (50%) because it was found based on pretesting that this reached indifference points faster: at short delays, most indifference points were relatively low, while at long delays indifference points were relatively high, and the 80% method allowed the staircase to reach the extremes of the scale more quickly.

Each staircase consisted of seven questions chosen in this manner, plus two questions to check for attention and/or railroading.[2] The first check was mean to test for consistency, and was chosen by taking the amount from the first question and adding or subtracting 2% to make an "easy" question. For example, if the participant initially chose $300 today over $1,709 in one year, the first check might be a choice between $300 today and $1,743. Clearly, the participant would be expected to choose the immediate $300 on the check question as well. The second

[2] Railroading would be if a participant made a mistake when answering the first question. If this happened, the subsequent questions would be unlikely to get near his/her indifference point.

check was meant to test whether the participant was always choosing the immediate option or always choosing the future option, without thinking. Therefore, the "correct" answer to the second check question was always designed to be the opposite of the answer given to the first question. This second check question posed the $300 immediate against an extremely large or small future amount, as appropriate (it was either the scale minimum divided by 2, or the scale maximum times 200). For example, if the participant initially chose $300 today over $1,709 in one year, the second check question might ask about $300 today or $20,000,000 in one year. On the other hand, if the participant initially chose $1,709 in the future, the second check might ask about $300 today versus $125 in the future.

Thus, each staircase consisted of nine questions total: seven regular questions, and two check questions. The check questions were the fifth and eighth questions, respectively. Pretesting indicated that participants enjoyed the check questions because they were easy to answer, giving them a break from the questions near their indifference points, which were difficult to answer.

The *multiple* part of the multiple staircase method came from the fact that three different scales were interleaved, one for each delay, in random order. So, participants were answering questions about 1-year, 10-year, and 50-year delays, in random order.

As a sample, here are the options that might be presented to one participant based on their choices. Note that each choice was presented one at a time, in contrast to the titration method, where all the options were presented on one page. The option the hypothetical participant chooses in each case is indicated with an X:

X Receive $300 now OR Receive $7,786 **fifty years from now**

Receive $300 now OR Receive $7,771 **one year from now** X

Receive $300 now OR Receive $7,737 **ten years from now** X

X Receive $300 now OR Receive $1,739 **ten years from now**

Receive $300 now OR Receive $1,764 **one year from now** X

X Receive $300 now OR Receive $82,087 **fifty years from now**

X Receive $300 now OR Receive $548 **one year from now**

Receive $300 now OR Receive Receive $6,574 **ten years from now** X

Receive $300 now OR Receive $96,620 **fifty years from now** X

X Receive $300 now OR Receive $2,690 **ten years from now**

X Receive $300 now OR Receive $311 **one year from now**

Receive $300 now OR Receive $85,257 **fifty years from now X**

X Receive $300 now OR Receive $5,747 **fifty years from now**

Receive $300 now OR Receive $9,688 **one year from now** X

Receive $300 now OR Receive $9,708 **ten years from now** X

Receive $300 now OR Receive $3,501 **ten years from now** X

X Receive $300 now OR Receive $356 **one year from now**

Receive $300 now OR Receive $84,733 **fifty years from now** X

Receive $300 now OR Receive $513 **one year from now** X

Receive $300 now OR Receive $82,385 **fifty years from now** X

Receive $300 now OR Receive $2,845 **ten years from now** X

X Receive $300 now OR Receive $125 **one year from now**

Receive $300 now OR Receive $20,176,000 **fifty years from now** X

X Receive $300 now OR Receive $126 **ten years from now**

X Receive $300 now OR Receive $2,832 **ten years from now**

Receive $300 now OR Receive $81,424 **fifty years from now** X

X Receive $300 now OR Receive $478 **one year from now**

The environmental multiple staircase was identical to the financial multiple staircase, but with two changes. The first was that there were four staircases (immediate, 1-year, 10-year, and 50-year) rather than three. The other was that the minimum amount was set to $0, based on pretesting which found that some participants placed a very low willingness-to-pay or willingness-to-accept for air quality.