Instructions for inferring discount rates from choices between immediate and delayed rewards

Kris N. Kirby
Williams College
October 2000

This document describes in detail a procedure for inferring values of the hyperbolic discount parameter (*k*) from a subject's choices between immediate and delayed rewards. The procedure is illustrated here by showing how to "score" the monetary-choice questionnaire used by Kirby, Petry, and Bickel (1999), which is shown in full at the end of the document. (This method was first used in Kirby & Marakovic, 1996.) This general procedure can be used to infer values of *k* in other contexts as well, such as for choices generated by the "staircase method" used by Rachlin, Green, and colleagues (see e.g., Green, Myerson, & Ostaszewski, 1999; Rachlin, Raineri, & Cross, 1991).

Data Entry

Enter choice responses into a spreadsheet with columns for each of the 27 questions on the questionnaire, and one row of data for each subject. Use a code of 0 for choices of the smaller, immediate reward (*SIR*) and 1 for choices of the larger, delayed reward (*LDR*).

It will be necessary to include a number of rows at the top of the spreadsheet that will allow the columns to be sorted properly for subsequent analyses. Examples are shown in Table 1, showing only the first seven questions on the questionnaire. The first row is labeled with the *LDR* size category. *S* indicates the smallest category (*LDRs* from $25 to $35 in row 4), *M* indicates the medium category (*LDRs* from $50 to $60 in row 4), and *L* indicates the largest category (*LDRs* from $75 to $85 in row 4). Because of the magnitude effect on discount rates (Kirby, 1997; Kirby & Marakovic, 1996; Kirby et al., 1999), it is important to estimate discount rates within each of the three reward size categories first, and then use the geometric mean of these three values if you need one overall value per person. The first row will be used to sort the questions into reward size categories.

The second row in Table 1 shows the question number, as it appeared on the questionnaire. This is useful to have if you need to sort the columns back into the order in

Table 1: Spreadsheet Layout of Example Data

	A	B	C	D	E	F	G	...
1	Size:	M	L	S	L	S	M	…
2	Quest:	1	2	3	4	5	6	…
3	SIR:	$54	$55	$19	$31	$14	$47	…
4	LDR:	$55	$75	$25	$85	$25	$50	…
5	Delay:	117	61	53	7	19	160	…
6	k:	0.00016	0.0060	0.0060	0.25	0.041	0.00040	…
7	S1	0	1	1	1	0	0	…
8	S2	0	1	0	1	1	0	…

Note: Size = the *LDR* size category, where *S* = smallest, *M* = medium, and *L* = largest; *Quest* = number of question in order on the questionnaire; *SIR* = smaller, immediate reward amount; *LDR* = larger, delayed reward amount; *Delay* is the delay to the *LDR* in days; *k* = hyperbolic discount parameter at indifference between the two rewards. Rows 7 and 8 contain the data for the first two subjects (S1 and S2): 0 = choice of *SIR*; 1 = choice of *LDR*.

which they were entered. The next three rows give the information about the rewards that were used on each trial. *Delay* is the delay to the *LDR*.

Using the value of the *SIR*, *LDR*, and the delay, one can calculate the value of *k* that would yield indifference between the *SIR* and *LDR*. Within reward-size categories one can use Mazur's (1987) equation, solved for *k*:

$$k = ((LDR/SIR)–1)/Delay \quad (1).$$

A person who had the value of *k* shown in row 6 of a given column should be indifferent between the two rewards in that column. For example, on question 2 in Table 1 (column C) subjects were asked to choose between \$55 today and \$75 in 61 days. Plugging into the equation, k = ((\$75/\$55)–1)/61 = 0.0060. So a person with a value of k = 0.0060 should be indifferent between \$55 today and \$75 in 61 days. The values of *k* that yield indifference on each question are shown in row 6. They will be used to sort the questions into descending order of *k* at indifference. The rewards that two subjects (S1 and S2) actually chose are shown in the bottom two rows of Table 1.

Sorting

After creating a spreadsheet with the column information shown in Table 1, you need to sort the 27 data columns of the spreadsheet so that the layout looks like Table 2. That is, you sort the 27 data columns by reward size (row 1), and then within reward size sort the nine columns by *k* (row 6) in *descending* order. This should tend to give you 1s changing to 0s as you move left to right across a subject's choices. (It takes a very impulsive subject to choose *SIR*s on the left in Table 2, and it takes a very unimpulsive subject to choose *LDR*s on the right in Table 2.) Table 2 only shows the values for the questions with *LDR*s in the "*L*" category. In the actual spreadsheet there would also be nine columns for the medium category and nine columns for the smallest category.

Table 2: Sorted Layout of Example Data

	A	B	C	D	E	F	G	H	I	J
1	Size:	L	L	L	L	L	L	L	L	L
2	Quest:	4	19	23	25	2	15	12	17	9
3	SIR:	\$31	\$33	\$41	\$54	\$55	\$69	\$67	\$80	\$78
4	LDR:	\$85	\$80	\$75	\$80	\$75	\$85	\$75	\$85	\$80
5	Delay:	7	14	20	30	61	91	119	157	162
6	k:	0.25	0.10	0.042	0.016	0.0060	0.0025	0.0010	0.00040	0.00016
7	S1	1	1	1	1	1	1	0	0	0
8	S2	1	1	0	1	1	0	0	0	0

Note: Size = the *LDR* size category, where *L* = largest; *Quest* = number of question in order on the questionnaire; *SIR* = smaller, immediate reward amount; *LDR* = larger, delayed reward amount; *Delay* is the delay to the *LDR* in days; *k* = hyperbolic discount parameter at indifference between the two rewards. Rows 7 and 8 contain the data for the first two subjects (S1 and S2): 0 = choice of *SIR*; 1 = choice of *LDR*.

Inferring Discount Rates from Choices

Subject 1 (S1), shown in row 7 of Table 2, chose the *LDR* on question 2 (now in column F: note the '1' in the cell F7). Either this subject preferred \$75 in 61 days over \$55 today, was indifferent between them and chose randomly, or made an error. Assuming that the choice was not an error, we may infer that this subject has a value of $k \leq 0.0060$. Had the subject chosen the *SIR* on this question, it would imply $k \geq 0.0060$. Each choice allows us to draw similar inferences, and the set of choices together allows us to put bounds on *k*.

Example 1. Consider again subject 1 (S1) in Table 2. Whenever a 1 appears in a column it implies a value of *k* less than the indifference value, whereas a 0 implies a value of *k* greater than the indifference value. S1 has 1s for each of the first six questions from left to right (columns B to G), together implying a value of $k \leq 0.0025$. This same subject has 0s for the last three questions (columns H to J), implying a value of $k \geq 0.0010$. Together this tells us that $0.0010 \leq k \leq 0.0025$. I take the geometric mean of the two values as the estimate of *k*, which in this case would be 0.0016. This is the value of *k* that I would record for this subject for the large reward-size category.

More generally, the nine values of *k* in Table 2 establish 10 ranges in which a person's actual value of *k* might fall. (I call these "bins.") Eight of these regions, those falling between any two data columns in Table 2, are bounded by the values of *k* on either side. For example, as we saw above, the interval between columns G and H is bounded by 0.0010 and 0.0025. I take the geometric midpoint of this interval to represent values of *k* that fall within the interval, in this case $k = 0.0016$.

The two regions beyond the endpoints, that is, to the left of column B or to the right of column J in Table 2, are unbounded. If a subject chooses all nine *SIR*s then we know her value of *k* is greater than 0.25, but we do not know by how much. If a subject chooses all nine *LDR*s then we know her value of *k* is less than 0.00016, but we do not know by how much. For those regions I simply use the values 0.25 and 0.00016, respectively, knowing that they are biased towards the middle. This shouldn't be a problem unless large proportions of subjects end up in these bins. The representative *k* values for all ten bins (eight bounded intervals and the values of the two endpoints) are shown in the row labeled k_{est} in Table 3.

Table 3: Regions ("bins") in which k *may fall*

	A	B	C	D	E	F	G	H	I	J
6	k_{ind}	.25	.10	.042	.016	.0060	.0025	.0010	.00040	.00016
7	S1	…								
8	S2	1	1	0	1	1	0	0	0	0

	bin_1	bin_2	bin_3	bin_4	bin_5	bin_6	bin_7	bin_8	bin_9	bin_{10}
k_{est}	.25	.159	.065	.026	.0098	.0039	.0016	.00063	.00025	.00016
con.	.56	.67	.78	.67	.78	.89	.78	.67	.56	.44

Note: k_{ind} = *k* at indifference; k_{est} = estimate of *k* for corresponding bin; *con*. = S2's consistency.

Handling Inconsistency

Example 2. Things get complicated when a subject's choices are not all consistent with a single value of *k*. Such a case is shown by subject 2 (see rows labeled "8" in Tables 2 and 3). This subject chose the *SIR* on question 23 (column D), which is inconsistent with the choices on either side of it in the table. One cannot simply look for a switch from 1s to 0s as one moves left to right across the table to estimate *k*.

I take a Bayesian approach to this problem, and find the value of *k* that was most likely to give rise to the observed set of choices. This requires considering each of the subject's choices in relation to each of the ten bins shown in Table 3. In essence, one asks what proportion of a subject's choices are consistent with each of the 10 assignable values of *k*. To illustrate, I will work through all 10 bins in Table 3. S2's choices are reprinted in the bottom row of Table 3.

(bin_1) Suppose we assume that S2 has a value of $k > 0.25$ (i.e., falls in the bin to the left of column B). If so, then this subject should have all zeros (i.e., should have always chosen the *SIR*). The subject actually chose the *SIR* on 5 of the 9 choices, so 5/9 = .56 of the subject's choices are consistent with assuming that her value of $k > 0.25$.

(bin_2) If we assume that S2 has a value of *k* = 0.159 (i.e., falls in the bin between columns B and C), then this subject should have chosen the *LDR* on question 4 (column B), and the *SIR* on the remaining eight trials (only zeros in columns C-J). The subject did choose the *LDR* on question 4, and chose the *SIR* on 5 of the remaining 8 trials. Therefore, (1+5)/9 = .67 of the subject's choices are consistent with assuming that her value of *k* = 0.159.

(bin_3) If we assume that S2 has a value of *k* = 0.065, then this subject should have chosen the *LDR* on questions 4 and 19 (columns B and C), and the *SIR* on the remaining seven trials (only zeros in columns D-J). The subject did choose the *LDR* on questions 4 and 19, and chose the *SIR* on 5 of the remaining 7 trials. Therefore, (2+5)/9 = .78 of the subject's choices are consistent with assuming that her value of *k* = 0.065.

(bin_4) If we assume that S2 has a value of *k* = 0.026, then this subject should have chosen the *LDR* on questions 4, 19 and 23 (columns B-D), and the *SIR* on the remaining six trials (only zeros in columns E-J). The subject chose the *LDR* on questions 4 and 19, but did not on question 23, and chose the *SIR* on 4 of the remaining 6 trials. Therefore, (2+4)/9 = .67 of the subject's choices are consistent with assuming that her value of *k* = 0.026.

(bin_5) If we assume that S2 has a value of *k* = 0.0098, then this subject should have chosen the *LDR* on questions 4, 19, 23, and 25 (columns B-E), and the *SIR* on the remaining five trials (only zeros in columns F-J). The subject chose the *LDR* on questions 4, 19 and 25, but did not on question 23, and chose the *SIR* on 4 of the remaining 5 trials. Therefore, (3+4)/9 = .78 of the subject's choices are consistent with assuming that her value of *k* = 0.0098.

(bin_6) If we assume that S2 has a value of *k* = 0.0039, then this subject should have chosen the *LDR* in columns B-F, and the SIR in columns G-J. The subject chose the *LDR* in 4 of the first 5 columns, and chose the *SIR* on all 4 of the remaining 4 trials. Therefore, (4+4)/9 = .89 of the subject's choices are consistent with assuming that her value of *k* = 0.0039.

(bin_7) If we assume that S2 has a value of *k* = 0.0016, then this subject should have chosen the *LDR* in columns B-G, and the *SIR* in columns H-J. The subject chose the *LDR* in 4 of the first 6 columns, and chose the *SIR* on 3 of the remaining 3 trials. Therefore, (4+3)/9 = .78 of the subject's choices are consistent with assuming that her value of *k* = 0.0016.

(bin_8) If we assume that S2 has a value of *k* = 0.00063, then this subject should have chosen the *LDR* in columns B-H, and the *SIR* in columns I and J. The subject chose the *LDR* in 4 of the first 7 columns, and chose the *SIR* on both of the two remaining trials. Therefore, (4+2)/9 = .67 of the subject's choices are consistent with assuming that her value of *k* = 0.00063.

(bin_9) If we assume that S2 has a value of *k* = 0.00025, then this subject should have chosen the *LDR* in columns B-I, and the *SIR* in column J. The subject chose the *LDR* in 4 of the first 8 columns, and chose the *SIR* in the last column. Therefore, (4+1)/9 = .56 of the subject's choices are consistent with assuming that her value of *k* = 0.00025.

(bin_{10}) If we assume that S2 has a value of $k < 0.00016$, then this subject should have chosen the *LDR* on all trials. The subject chose the *LDR* on 4 of the trials. Therefore, 4/9 = .44 of the subject's choices are consistent with assuming that her value of $k < 0.00016$.

As we have seen, the maximum consistency value was obtained by assuming (bin_6) that S2 has a value of $k = 0.0039$. This is the value that yields the highest consistency among the subject's choices, and is, in that sense, the value that was most likely to have given rise to the observed choices. This is the value that I would assign for S2 for the large reward-size category.

Computation in *Microsoft Excel*

I use *Excel* to compute all of the consistency numbers. First, I create ten new columns corresponding to the ten bins in Table 3. (I insert the ten new columns, and all additional columns for the *L* data, in between the data columns for the *L* and *M* reward-sizes.) Example columns are shown in Table 4, with *k*s representing each of the ten bins shown in row 2. (I also insert empty columns for labels and to separate the columns, which is why the data begins in column M in Table 4.)

Then I calculate the proportion of choices consistent with each of those ten bins. The general formula for this computation is

consistency = (#correctly predicted *LDR* choices + #correctly predicted *SIR* choices) / 9

For example, when assigning S2 to bin_6 we predicted 5 *LDR* choices in columns B-F, and observed 4 such choices. We predicted 4 *SIR* choices in columns G-J and observed all 4. This gives

consistency = (4 + 4) / 9 = .89.

To do this in *Excel* using the 0/1 coding described above, find the number of correctly predicted *LDR* choices by summing the 1s in columns that are predicted to have 1s. Find the number of correctly predicted *SIR* choices by summing the 1s in columns expected to have 0s, and subtracting this value from the number of expected 0s. In our example, for bin_6 this gives

consistency = [Σ(columns B-F) + 4 – Σ(columns G-J)] / 9 = [4 + (4 – 0)] / 9 = .89.

The "4" between the plus and minus signs is the number of zeros expected in columns G-J. This value is contained in a cell in the same column as the bin to which it corresponds, as shown in row 1 of Table 4. The k_{est} for bin_6 is in column R, so the expected number of 0s is in R1.

Table 4: The ten bins in which k *may fall, and their representative* k *values*

	L	M	N	O	P	Q	R	S	T	U	V
1	E(0s):	9	8	7	6	5	4	3	2	1	0
2	k_{est}:	.25	.159	.065	.026	.0098	.0039	.0016	.00063	.00025	.00016
...											
8	Cons:	.56	.67	.78	.67	.78	.89	.78	.67	.56	.44

Note: E(0s) = number of expected 0s; k_{est} = estimate of *k* for corresponding bin; *con.* = S2's consistency.

Using a combination of absolute and relative references, one can write a formula to calculate the consistencies and paste it into new cells, without having to write a different formula for each cell. For example, one might type into cell R8 in Table 4 the following formula (using *Excel* notation):

= (SUM($B8:F8) + R$1 – SUM(G8:$J8)) /9

Absolute references are denoted by dollar signs ($). In Table 2, the columns containing the nine *L* questions are always in columns B to J, and the $B and $J references will remain constant even when they are pasted into new cells. Likewise, the expected number of 1s is always in row 1, which is kept fixed by the $1 reference. If this formula is now copied and pasted into cell S8 in Table 4, for example, all of the references that are not preceded by $ signs will be updated to retain their relative position to the new cell (i.e., they will be shifted over by one row). The pasted formula becomes:

= (SUM($B8:G8) + S$1 – SUM(H8:$J8)) /9

This formula can be pasted into all cells in columns N through U for all subjects.

To handle the consistencies for the endpoints (bins 1 and 10), the formula needs to be changed. For the highest value of *k*, that is, to find the consistency value in cell M8 in Table 4, paste in the full formula and then delete the first term in the numerator (the one that sums up *LDR* choices). The formula for this cell (M8) should then be:

= (M$1 – SUM(B8:$J8)) /9

(One could add a $ before the B, but because this formula is never pasted into a different column this makes no difference.) This formula can now be pasted down column M for all subjects.

For the smallest value of *k*, that is, to find the consistency value in cell V8 in Table 4, paste in the full formula (for example, by copying from R8 and pasting into V8) and then delete the second and third terms in the numerator (the ones that give the number of consistent *SIR* choices). The formula for this cell (V8) should be:

= SUM($B8:J8) /9

This formula can now be pasted down column V.

After generating all of the consistency values, I create a new column, say column W, that finds the maximum consistency for each row. For row 8, the formula in W8 would be

= MAX($M8:$V8)

Picking Out *k*s That Yield Maximum Consistency

The final step is to pick out those values of *k* that correspond to the maximum consistency among choices. Sometimes this is a single value, but there can be ties. I handle ties by finding the geometric mean of all *k* values that correspond to the maximum level of consistency for the given subject.

To pick out the *k* values that correspond to the maximum level of consistency, create ten new columns, say columns Y-AH (actually, 11, leaving column X blank as a visual separator). Each new column compares one of the columns M through V with the maximum consistency, which I am assuming is contained in column W. If the consistency for a given column is equal to the maximum, set the new cell equal to the corresponding *k* value, which recall is in row 2 of the corresponding column M-V (see Table 4). Otherwise the cell is set to missing (a period). Here is the formula for the first column (say, column Y), the one checking the consistency for the highest value of *k*:

=IF(M8=$W8, M$2, ".")

In words, if the consistency in cell M8 is equal to the maximum consistency for that subject shown in cell W8, then set this cell (Y8) equal to cell M2, which contains the corresponding value of *k*. Otherwise, set the cell (Y8) equal to ".". This formula can be pasted over all ten columns and down through all subjects.

Each row of these ten new columns will contain all of the *k* values for a given subject that yield the maximum level of consistency among choices. For S2 only one bin (6) has the maximum level of consistency, so all of the columns Y-AH would have "." in row 8, except for the sixth column (AD8) which would contain the value 0.0039.

Finally, create one new column, say in column AI, that finds the geometric mean across the ten columns of *k* values. When there is only one value of *k* corresponding to the maximum consistency among choices this will simply return that value. When two or more bins are tied at maximum consistency, this will return the geometric mean of their values. In *Excel* the formula for the geometric mean for cell AI8 would be:

=GEOMEAN(Y8:AH8)

This formula can be pasted all the way down the column, and will give the estimate of *k* that is most likely to have given rise to the observed choices. This procedure can be repeated for the small and medium delayed reward sizes in a similar manner.

References

Green, L., Myerson, J., & Ostaszewski, P. (1999). Amount of reward has opposite effects on the discounting of delayed and probabilistic outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 418-427.

Kirby, K. N. (1997). Bidding on the future: Evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General, 126*(1), 54-70.

Kirby, K. N., & Marakovic, N. N. (1996). Delay-discounting probabilistic rewards: Rates decrease as amounts increase. *Psychonomic Bulletin and Review, 3*(1), 100-104.

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General, 128*(1), 78-87.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons & J. E. Mazur & J. A. Nevin & H. Rachlin (Eds.), *Quantitative analyses of behavior, Vol. 5: The effect of delay and of intervening events on reinforcement value* (Vol. 5, pp. 55-73). Hillsdale, NJ: Erlbaum.

Rachlin, H., Raineri, A., & Cross, D. (1991). Subjective probability and delay. *Journal of the Experimental Analysis of Behavior, 55*(2), 233-244.

Acknowledgements
I thank Cornelius Koch for helpful suggestions on a previous draft. Preparation of this document was supported by grant MH55262 from the National Institute of Mental Health. Please send suggestions or corrections to kkirby@williams.edu.

Monetary-Choice Questionnaire

For each of the next 27 choices, please indicate which reward you would prefer: the smaller reward <u>today</u>, or the larger reward in the specified number of days.

Please take the choices seriously: they are for <u>REAL MONEY</u>. At the end of the session one of the 27 questions will be selected at random and you will get the reward that you chose on that question. If you choose the smaller reward, you will get paid before you leave today. If you choose the delayed reward, you will get paid in the specified number of days and not before. So to make sure that you get a reward you prefer, you should answer <u>every question</u> as though it were the one you will win.

1. Would you prefer $54 today, or $55 in 117 days?
2. Would you prefer $55 today, or $75 in 61 days?
3. Would you prefer $19 today, or $25 in 53 days?
4. Would you prefer $31 today, or $85 in 7 days?
5. Would you prefer $14 today, or $25 in 19 days?
6. Would you prefer $47 today, or $50 in 160 days?
7. Would you prefer $15 today, or $35 in 13 days?
8. Would you prefer $25 today, or $60 in 14 days?
9. Would you prefer $78 today, or $80 in 162 days?
10. Would you prefer $40 today, or $55 in 62 days?
11. Would you prefer $11 today, or $30 in 7 days?
12. Would you prefer $67 today, or $75 in 119 days?
13. Would you prefer $34 today, or $35 in 186 days?
14. Would you prefer $27 today, or $50 in 21 days?
15. Would you prefer $69 today, or $85 in 91 days?
16. Would you prefer $49 today, or $60 in 89 days?
17. Would you prefer $80 today, or $85 in 157 days?
18. Would you prefer $24 today, or $35 in 29 days?
19. Would you prefer $33 today, or $80 in 14 days?
20. Would you prefer $28 today, or $30 in 179 days?
21. Would you prefer $34 today, or $50 in 30 days?
22. Would you prefer $25 today, or $30 in 80 days?
23. Would you prefer $41 today, or $75 in 20 days?
24. Would you prefer $54 today, or $60 in 111 days?
25. Would you prefer $54 today, or $80 in 30 days?
26. Would you prefer $22 today, or $25 in 136 days?
27. Would you prefer $20 today, or $55 in 7 days?

ID: ________________________ *Gender:* male / female *Age:* ____

How you can be reached for reward delivery: