Online Appendix for *Additional deliberation reduces pessimism: evidence from the double-response method*

Appendix A: Experimental Instructions (DR condition)

Thank you for participating in this experiment! Just for being here on time you will earn 5 PLN. You can keep this amount regardless of the outcome of the experiment. Any further payoff will depend on how much you earn during the experiment, in accordance with the procedure specified in these instructions.

INSTRUCTIONS

In today's experiment you will make several decisions over a number of rounds. In each round will see an urn with eight balls. The balls may have different colours: black, blue, green, yellow, pink, red, brown and turquoise.

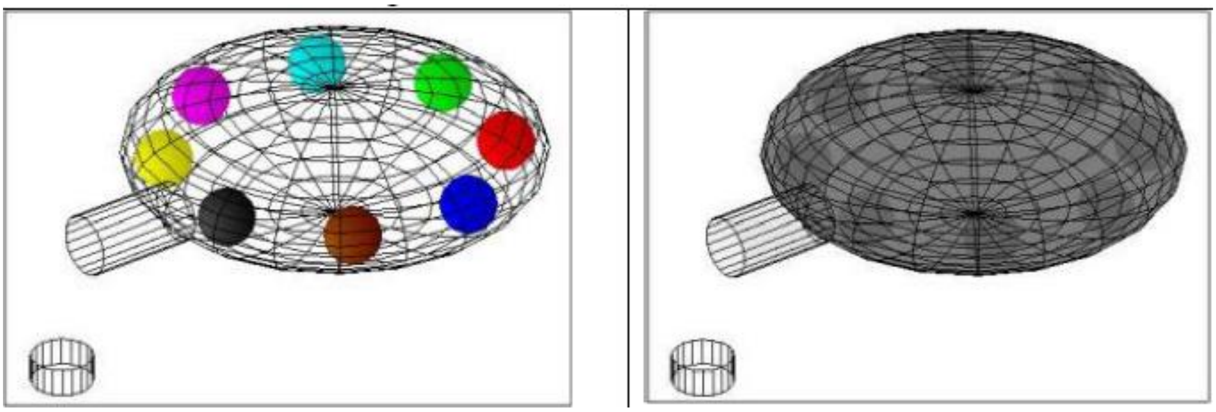

Fig. 1a. TRANSPARENT URN
In an urn like this you can see one ball in each of the eight colours

Fig. 1b. OPAQUE URN
With an urn like that you cannot know how many balls of each colour are inside

Figure 1 shows two kinds of urns that you can face. **Figure 1a. shows a transparent urn**. It will always contain exactly eight balls, with one ball in each of eight colours. This means that **the probability of selecting each colour is the same. The opaque urn shown in Figure 1b**. also has exactly eight balls. However, you cannot see the colours of the balls in this urn. **This means that you cannot know exactly how many balls of each colour are there.** It may happen that some of these colours are missing, while others will show up more than once. You will not know the exact composition of the urn. The probabilities of selecting different colours cannot be known and may be different for different colours.

We are interested in finding out how much each of a number of lotteries represented by urns is worth to you. You will be asked to make two decisions: an initial decision and a final decision in each of 34 rounds (including two trial rounds). In each of them you will have to indicate how much a lottery is presently worth to you. You will be asked to type in this amount in the dedicated field on the screen, as illustrated in Figure 2. Please note the possible payoffs and the number of balls resulting in each possible outcome. Both the higher and lower amount that may result from the lottery an the number of balls resulting in these outcomes will vary across decision tasks. Because your final monetary payoff will depend on these decisions, you should carefully analyse these aspects of the choices you make. Figure 2 shows an example of a decision task involving a transparent urn. At the top you can see how much time is left till the end of the round, which round it is and which phase of the run. Below you can see a transparent or opaque urn and information on how many colours are associated with winning each specific amount. Below you can see the question “How much is this lottery worth to you?”, prompting an answer to be typed in and confirmed.

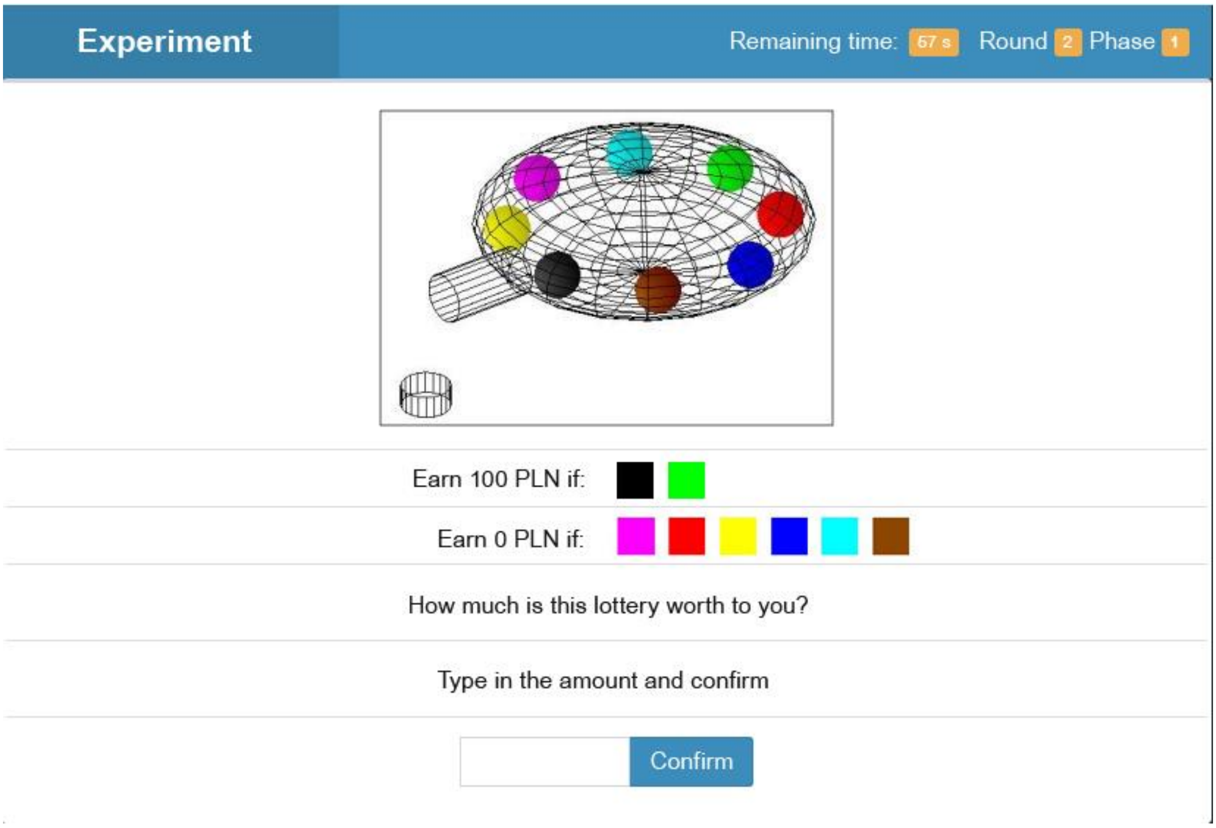

Fig. 2: An example of a decision task.

NOTE: unlike in most experiments in our Laboratory, TIME will play a very important role in today’s experiment. Each round will last up to 60 seconds and <u>within this time you will have to make two decisions: initial and final.</u> Ideally, you would like to keep indicating, at any moment of the round, what you currently consider to the best choice. Try to EVALUATE THE LOTTERY AS FAST AS YOU CAN AND ENTER YOUR INITIAL DECISION by typing in the amount and clicking “confirm” or pressing the Enter button. Then go back to the description of the lottery and think again. IF YOU CHANGE YOUR MIND, CHANGE YOUR CHOICE ON THE SCREEN ACCORDINGLY. Type in a new amount and click “confirm” or press Enter. You can also leave your initial choice unchanged: IF YOU COME TO A CONCLUSION THAT THE INITIALLY TYPED IN AMOUNT IS OPTIMAL, RE-TYPE IT ONCE MORE and click “confirm” or press Enter. Upon confirming the final decision you will be prompted to move to the next round by clicking the “next round” button.

To encourage possibly quick but at the same time careful consideration we will use the following method to determine your payoffs. At the end of the experiment the computer will not only randomly choose one round to determine your payoffs, but also a SPECIFIC SECOND of this round. The choice that was indicated by you at this specific second in this round will be implemented. If the computer chooses a second, in which you had not managed to choose any option yet, one of the options will be chosen randomly. Typically, it will be less profitable for you than have your own, conscious choices implemented. It means that it is best to make your initial decision very quickly (but not too quickly, it would effectively be random again in such a case), whereas if you realize that the initial choice was not optimal, to type in and confirm the modified amount.

We always randomly pick one of the 60 seconds, no matter how long the round actually lasted. If we pick on the of the seconds after your final decisions, this decision will be implemented.

Example

In round 9, a participant was evaluating a lottery represented by a transparent urn (with known probabilities). It involved winning 100 PLN with probability 4/8 and 40 PLN otherwise (thus also with probability 4/8). The participant initially assessed that the lottery is worth 60 PLN to her. She typed in this amount and confirmed it in the 10th second of the round. She knew, however, that the initial decision may not be optimal and that the payoff may be determined by the decision made at some later second of the round. She thus looked at the lottery again and realized that it is worth more to her than she initially thought. She thus eventually changed the decision (in the 44th second of the round), by typing in a new amount, 72 PLN and clicking "confirm", thereby ending the round.

Let us now assume that at the end of the experiment the computer randomly picked round 9. Simultaneously, a specific second of this round (1-60) is selected. Let us assume for example, that the 15th second is selected. Thus the initial decision, confirmed in the 10th second, is implemented. We thus understand that the decision maker evaluated the lottery at 60 PLN.

The computer randomly picks a number from the range between the lowest and the highest payoff in the lottery (here: 40-1000) this number can be interpreted as the amount offered to the participant instead of the lottery. If this amount is higher than the signalled value of the lottery, she will receive this amount. If it is lower – she will receive the lottery. Assume for example that the amount of 48 PLN is selected. The participant likes the lottery (evaluated at 60 PLN) better than this amount. Thus the lottery will be played: the participant will get 100 with probability 4/8 and 40 PLN otherwise. By contrast, if the randomly picked number was higher than her evaluation of the lottery (equal to 60 PLN), for example equal to 70 PLN, the participant will receive this amount instead of running the lottery.

If a later second of the round is picked, one by which the participant has managed to confirm her final decision, for example the 47th second of the round, this final decision will be implemented. For example, if the randomly picked sure amount offered instead of the lottery is 70 PLN as before, this time it will be lower than the participant's evaluation of this lottery (72 PLN). This time, instead of getting 70 PLN for sure, the participant will play the lottery.

By contrast, if one of the first 9 seconds (in which no choice was made yet) is picked, the computer will randomly pick the lottery or the randomly picked amount being offered instead of the lottery.

Even if the mechanism described above seems complicated, its consequences are simple: it is in your best interest to make a quick initial decision and modify your decision as soon as you come to a conclusion that the initial decision was not optimal. If you strengthen your belief that it was indeed optimal, you can re-type it and confirm to move on to the subsequent round.

Appendix B: Cognitive Reflection Test

Question 1: A bat and a ball cost 110 PLN in total. The bat costs 100 more than the ball. How much does the ball cost? (correct answer: 5. intuitive: 10).

Question 2: In a lake. there is a patch of lily pads. Every day. the patch doubles in size. If it takes 48 days for the patch to cover the entire lake. how long would it take for the patch to cover half of the lake? (correct answer: 47. intuitive: 24).

Question 3: If it takes 5 machines 5 minutes to make 5 widgets. how long would it take 100 machines to make 100 widgets? (correct answer: 5. intuitive: 100).

The scores on the Cognitive Reflection Test (CRT) turned out to be strongly correlated with choices (unlike other – demographic – items of the post-experiment questionnaire). Specifically, those with low score on the CRT tended to provide higher CEs. This tendency was highly significant initial and final choices in all conditions. No link with the number or direction of changes, nor decision times was found.

Appendix C: Supplementary tables

Table C1 shows mean certainty equivalents in all the treatments. In the Double Response Treatment, differences between the initial and the final decisions were very small, but usually the latter were less risk-averse (showed higher certainty equivalents). Overall, the initial decisions were modified downwards in 7% of the cases, upwards in 16% of the cases and left unchanged in 77%. For 18 out of 32 rounds (problems number 3, 5, 6, 7, 11, 13, 14, 17, 18, 19, 20, 21, 22, 24, 28, 30, 31, and 32) these differences were significant at 10% (and for 14 of these at 5%) in a Wilcoxon test, see Table C2. Clearly, that is much more than 3.2 (and 1.6 respectively) significant differences that would be expected to arise by pure chance.

Interestingly, it seems that the Double Response method is indeed crucially important in identifying such a moderate effect. Specifically, we run the following exercise: for a moment we disregard the fact that we have two CEs (initial and final) from the same individual in each decision problem and instead we treat them as if they were coming from separate individuals. We then test using Mann-Whitney U test (based on unpaired data) for differences between initial and final choices in each decision problem separately. The tests statistics are within 5% region in 1 out of 32 rounds. This suggests that without the possibility to look at within-subject differences one would miss the effect of time pressure that is identified here.

Comparing DR against NTP, final CEs were significantly different only in 7 out of 32 cases at 10% level (problems number 8, 19, 20, 21, 22, 23, and 29) and only in 1 out of 32 cases at 5% level in a Mann-Whitney U test, see Table C3. These figures are rather comparable to the null-hypothesis benchmarks of 3.2 and 1.6.

To summarize, reconsideration after longer deliberation period tended to make participants a bit less risk averse and their final decisions were, on average, quite similar to those made under no time pressure at all. In the next subsection we show how these tendencies translate into estimated probability weighting functions.

Table C1. Mean certainty equivalents by treatment

			Double Response (n=113)				No Time Pressure (n=38)	
	URN	LOTTERY	MEAN Initial decision	SD Initial decision	MEAN Final decision	SD Final decision	MEAN	SD
1	known	$\frac{1}{8}$ 100PLN; $\frac{7}{8}$ 0PLN	31.58	28.28	31.17	27.99	28.79	25.86
2	known	$\frac{2}{8}$ 100PLN; $\frac{6}{8}$ 0PLN	36.50	23.34	36.37	22.59	35.68	21.72
3	known	$\frac{3}{8}$ 100PLN; $\frac{4}{8}$ 0PLN	44.46	19.69	45.10	19.85	44.50	20.93
4	known	$\frac{4}{8}$ 100PLN; $\frac{4}{8}$ 0PLN	53.73	19.26	54.39	19.57	53.39	21.06
5	known	$\frac{5}{8}$ 100PLN; $\frac{3}{8}$ 0PLN	62.81	19.39	64.16	19.13	61.68	21.19
6	known	$\frac{6}{8}$ 100PLN; $\frac{2}{8}$ 0PLN	73.25	16.19	74.99	16.40	73.42	19.99
7	known	$\frac{7}{8}$ 100PLN; $\frac{1}{8}$ 0PLN	84.92	16.74	86.35	15.29	82.68	20.24
8	known	$\frac{4}{8}$ 60PLN; $\frac{4}{8}$ 0PLN	35.92	12.56	36.28	12.59	32.18	12.06
9	known	$\frac{4}{8}$ 100PLN; $\frac{4}{8}$ 40PLN	65.97	14.12	66.42	13.21	64.97	12.93
10	known	$\frac{4}{8}$ 40PLN; $\frac{4}{8}$ 0PLN	23.00	9.34	23.12	9.05	22.00	9.33
11	known	$\frac{4}{8}$ 60PLN; $\frac{4}{8}$ 20PLN	40.12	10.44	41.04	9.84	38.95	9.51
12	known	$\frac{4}{8}$ 80PLN; $\frac{4}{8}$ 40PLN	59.88	10.42	59.95	9.98	59.66	11.07
13	known	$\frac{4}{8}$ 100PLN; $\frac{4}{8}$ 60PLN	77.65	10.42	79.16	10.05	79.11	9.92
14	unknown	$\frac{1}{8}$100PLN; $\frac{7}{8}$ 0PLN	25.88	23.37	27.34	23.39	25.26	25.03
15	unknown	$\frac{1}{8}$100PLN; $\frac{7}{8}$ 0PLN	30.66	25.79	31.04	25.92	25.08	23.56
16	unknown	$\frac{1}{8}$100PLN; $\frac{7}{8}$ 0PLN	33.42	27.85	34.02	27.20	29.00	27.76
17	unknown	$\frac{2}{8}$ 100PLN; $\frac{6}{8}$ 0PLN	38.70	23.51	39.93	26.61	33.84	25.80
18	unknown	$\frac{2}{8}$ 100PLN; $\frac{6}{8}$ 0PLN	36.72	21.81	37.95	22.28	31.13	20.56
19	unknown	$\frac{2}{8}$ 100PLN; $\frac{6}{8}$ 0PLN	40.36	24.70	41.11	25.39	32.61	22.84
20	unknown	$\frac{2}{8}$ 100PLN; $\frac{6}{8}$ 0PLN	38.86	23.20	41.03	24.46	32.63	24.24
21	unknown	$\frac{3}{8}$ 100PLN; $\frac{5}{8}$ 0PLN	47.97	24.36	49.55	24.28	41.03	22.86

22	unknown	$\frac{4}{8}$ 100PLN; $\frac{4}{8}$ 0PLN	52.52	23.19	54.05	22.98	45.42	23.02
23	unknown	$\frac{4}{8}$ 100PLN; $\frac{4}{8}$ 0PLN	53.37	21.575	54.14	20.71	45.34	23.58
24	unknown	$\frac{5}{8}$ 100PLN; $\frac{3}{8}$ 0PLN	62.32	21.45	63.22	22.33	55.26	24.17
25	unknown	$\frac{6}{8}$100PLN; $\frac{2}{8}$ 0PLN	69.51	22.77	70.53	22.06	63.45	25.72
26	unknown	$\frac{7}{8}$ 100PLN; $\frac{1}{8}$ 0PLN	78.72	23.49	79.96	21.94	73.82	26.98
27	unknown	$\frac{4}{8}$ 60PLN; $\frac{4}{8}$ 0PLN	36.12	14.94	35.94	15.01	31.39	16.49
28	unknown	$\frac{4}{8}$ 100PLN; $\frac{4}{8}$ 40PLN	61.88	14.07	62.95	13.47	62.16	12.13
29	unknown	$\frac{4}{8}$ 40PLN; $\frac{4}{8}$ 0PLN	23.60	10.90	23.89	10.58	20.29	10.15
30	unknown	$\frac{4}{8}$ 60PLN; $\frac{4}{8}$ 20PLN	38.23	10.34	39.39	9.56	37.03	9.17
31	unknown	$\frac{4}{8}$ 80PLN; $\frac{4}{8}$ 40PLN	58.45	10.10	59.12	9.36	57.00	10.12
32	unknown	$\frac{4}{8}$ 100PLN; $\frac{4}{8}$ 60PLN	76.67	10.48	78.09	10.20	77.66	9.58

Table C2. Wilcoxon tests for equality of final vs. initial Certainty Equivalents in DR

Round	***z***	***p***
1	-0.42	0.675
2	-0.272	0.786
3	-1.736	0.083
4	-.829	0.407
5	-2.270	0.023
6	-2.810	0.005
7	-2.633	0.008
8	-1.219	0.223
9	-1.094	0.274
10	-.515	0.607
11	-1.819	0.069
12	-.789	0.430
13	-2.824	0.005
14	-2.096	0.036
15	-.818	0.413
16	-.804	0.421

Round	***z***	***p***
17	-2.302	0.021
18	-1.793	0.073
19	-2.058	0.040
20	-3.051	0.002
21	-2.852	0.004
22	-2.120	0.034
23	-1.453	0.146
24	-2.609	0.009
25	-1.313	0.189
26	-1.309	0.191
27	-.473	0.636
28	-2.277	0.023
29	-.874	0.382
30	-3.018	0.003
31	-2.172	0.030
32	-2.707	0.007

Table C3. Mann Whitney U tests for equality of final Certainty Equivalents: DR vs. NTP

ROUND	***z***	***p***
1	-0.506	0.613
2	-0.078	0.938
3	-0.245	0.806
4	-0.072	0.943
5	-0.388	0.698
6	-0.048	0.962
7	-0.953	0.341
8	-1.705	0.088
9	-0.683	0.494
10	-0.678	0.498
11	-1.615	0.106
12	-0.250	0.803
13	-0.245	0.807
14	-0.906	0.365
15	-1.542	0.123
16	-1.449	0.147

ROUND	***z***	***p***
17	-1.725	0.084
18	-1.677	0.094
19	-1.875	0.061
20	-2.143	0.032
21	-1.796	0.072
22	-1.912	0.056
23	-1.962	0.050
24	-1.551	0.121
25	-1.445	0.149
26	-1.013	0.311
27	-1.512	0.130
28	-0.041	0.967
29	-1.801	0.072
30	-1.181	0.238
31	-0.861	0.389
32	-0.077	0.939

Table C4. Spearman correlation of some individual variables.

		Final_ak	Initial_au	Final_au
Initial_ak	*r*	.822**	.397**	.396**
	p	.000	.000	.000
Final_ak	*r*		.367**	.376**
	p		.000	.000
Initial_au	*r*			.939**
	p			.000

		Final_bk	Initial_bu	Final_bu
Initial_bk	*r*	.835**	.529**	.474**

	p	.000	.000	.000
Final_bk	*r*		.554**	.576**
	p		.000	.000
Initial_bu	*r*			.915**
	p			.000

Appendix D: Individual behavior

Our design allows analyzing each participant's choices separately. In particular, it is of interest how participant-level parameters correlate with each other. Table D1a shows, not surprisingly, that individual α parameters estimated using initial and final choices in known urns are strongly correlated. Likewise, αs based on unknown urns are highly consistent. Correlations across domains are much lower but still substantial: participants that are sensitive to probabilities for known urns tend to be also sensitive to probabilities for unknown urns.

Table D1a. Spearman correlations for measures of sensitivity to probability

		Final α (K)	Initial α (U)	Final α (U)
Initial α (K)	*r*	.817**	.419**	.449**
	p	.000	.000	.000
Final α (K)	*r*		.330**	.367**
	p		.000	.000
Initial α (U)	*r*			.926**
	p			.000

Table D1b. Spearman correlations for measures of pessimism

		Final β (K)	Initial β (U)	Final β (U)
Initial β (K)	*r*	.831**	.489**	.480**
	p	.000	.000	.000
Final β (K)	*r*		.489**	.515**
	p		.000	.000
Initial β (U)	*r*			.861**
	p			.000

The picture is rather similar for the βs, which are highly consistent within domain (across deliberation times) and moderately correlated across domains (K vs. U), see Table D1b. That is to say, participants who are pessimistic when facing known urns tend to be pessimistic for unknown urns as well.

We may also classify participants based on their choices. Because we do not have a very specific theoretical benchmark, we use a "theory-free" approach, namely cluster analysis. We applied the K-means method separately for known and unknown urns.[1] Note that we have based this classification on initial choices only. This allows inferring if, say, the initially

[1] The AIC and BIC criteria did not provide a consistent answer as to the optimal number of clusters. Here we report the case of three clusters, which, for the known urns, seems to deliver a naturally interpretable division.

pessimistic group changes its behavior more after deliberation than the groups which is rational in the first place. The resulting sizes of the clusters are provided in Table D2.

Table D2. Number of participants in each cluster

Urn	**Cluster**	**Number of participants**
Known	1	53
	2	24
	3	36
Unknown	1	35
	2	30
	3	48

Table D3 presents median probability weights for known and unknown urns; the resulting values of α and β estimated from choices on known urns are graphically represented in Figures D1-D3 and those for unknown urns in Figures D4-D6.

Table D3. Median Probability weights by cluster and treatment.

			P						
Urns	treatment	cluster	**.125**	**.250**	**.375**	**.500**	**.625**	**.750**	**.875**
K	DR:Initial	**1**	.114	.239	.360	.493	.630	.750	.901
	DR:Final		.120	.250	.370	.478	.636	.758	.906
	DR:Initial	**2**	.694	.474	.371	.341	.563	.543	.669
	DR:Final		.696	.383	.444	.419	.587	.623	.789
	DR:Initial	**3**	.106	.173	.303	.408	.504	.621	.801
	DR:Final		.110	.226	.304	.469	.533	.672	.813
U	DR:Initial	**1**	.084	.140	.222	.310	.429	.495	.701
	DR:Final		.110	.162	.230	.381	.443	.543	.701
	DR:Initial	**2**	.197	.303	.536	.313	.587	.703	.846
	DR:Final		.215	.391	.546	.355	.590	.732	.916
	DR:Initial	**3**	.084	.141	.314	.390	.535	.676	.827
	DR:Final		.081	.146	.329	.400	.539	.689	.858

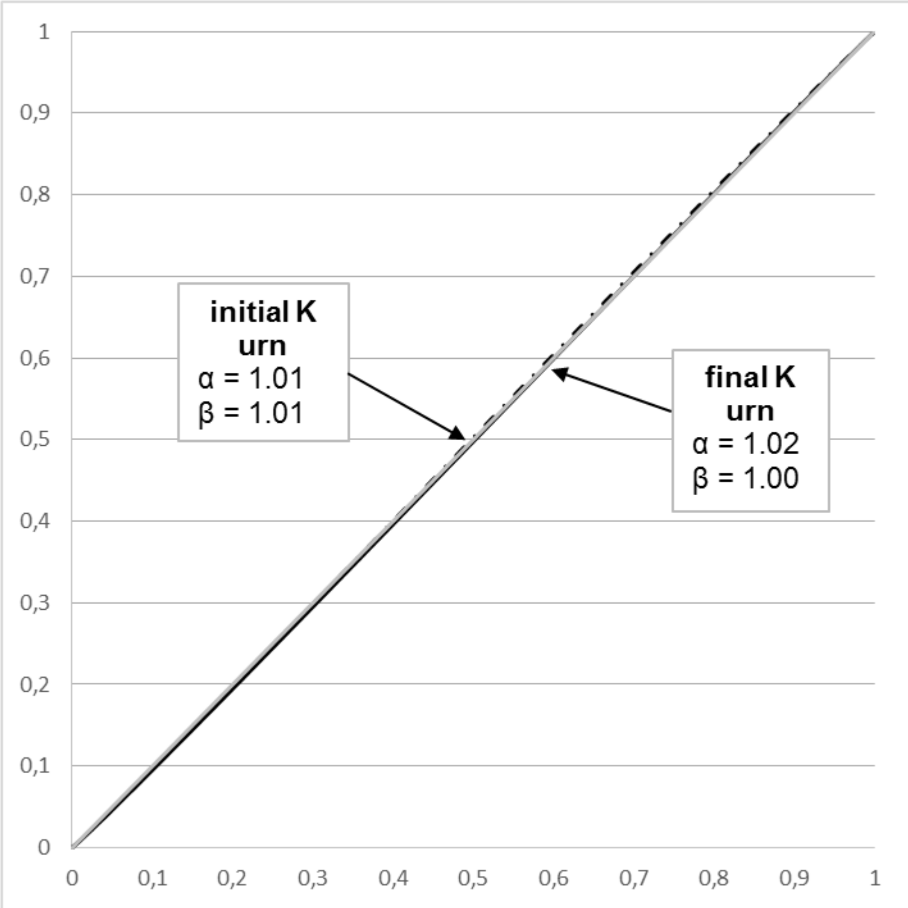

Fig. D1. Median individual probability weighting functions: known urns – Cluster 1

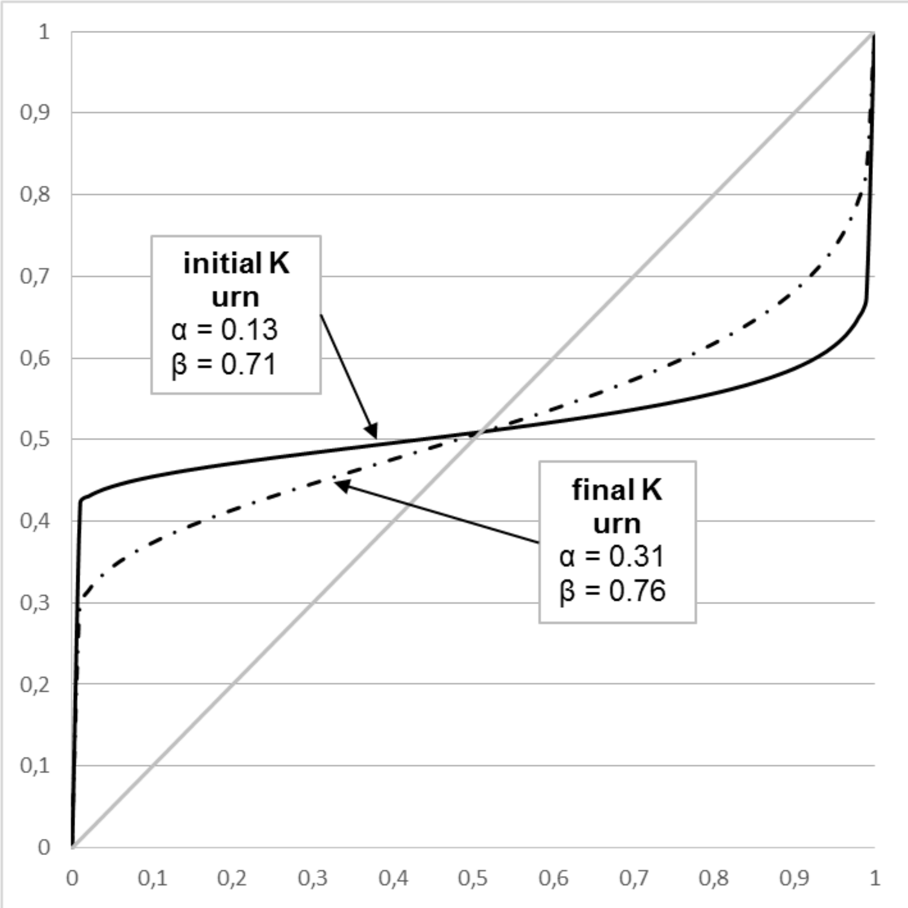

Fig. D2. Median individual probability weighting functions: known urns – Cluster 2

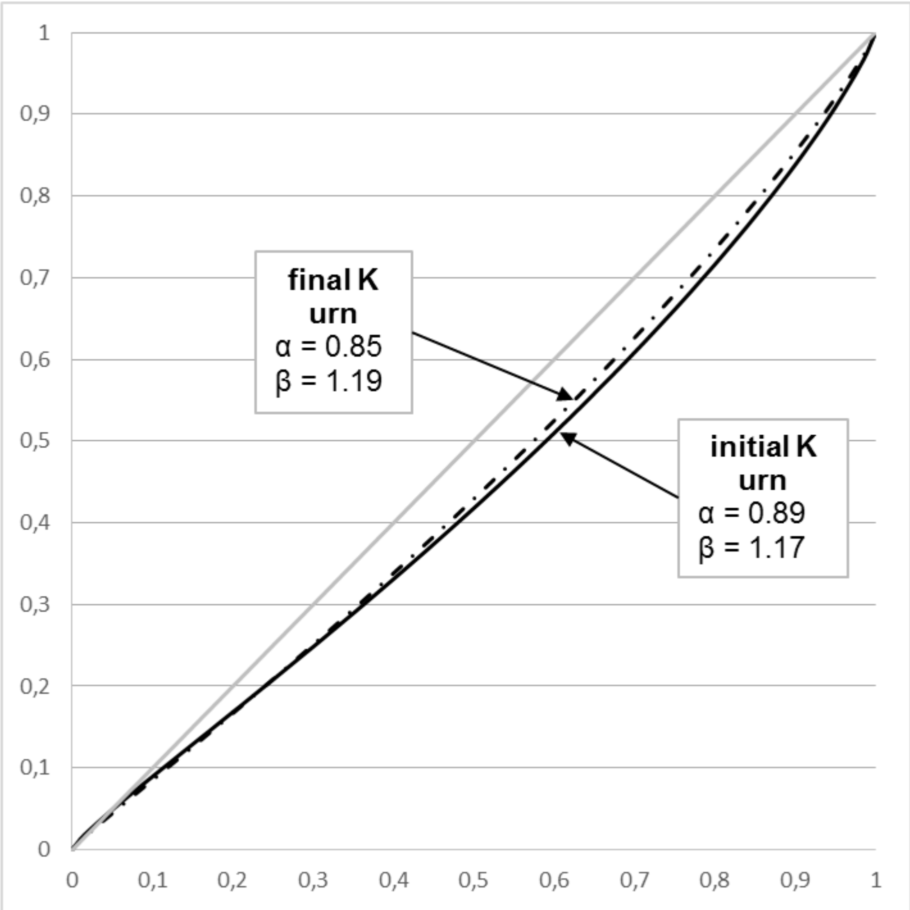

Fig. D3. Median individual probability weighting functions: known urns – Cluster 3

As can be seen, three distinct patterns can be seen in the case of known urns. The largest fraction of participants (Cluster 1) behave consistently with the rational model of linear probability weighting. Participants in Cluster 2 show a strong inverse-S pattern, whereby they overweight small probabilities of success and underweight large probabilities. Finally, those in Cluster 3 are simply pessimistic, consistently behaving as if the probability of success was lower than provided. Interestingly, there was some tendency to adjust the weights upwards after deliberation in each of these clusters, albeit, predictably, it was most pronounced in the "pessimistic" Cluster 3.

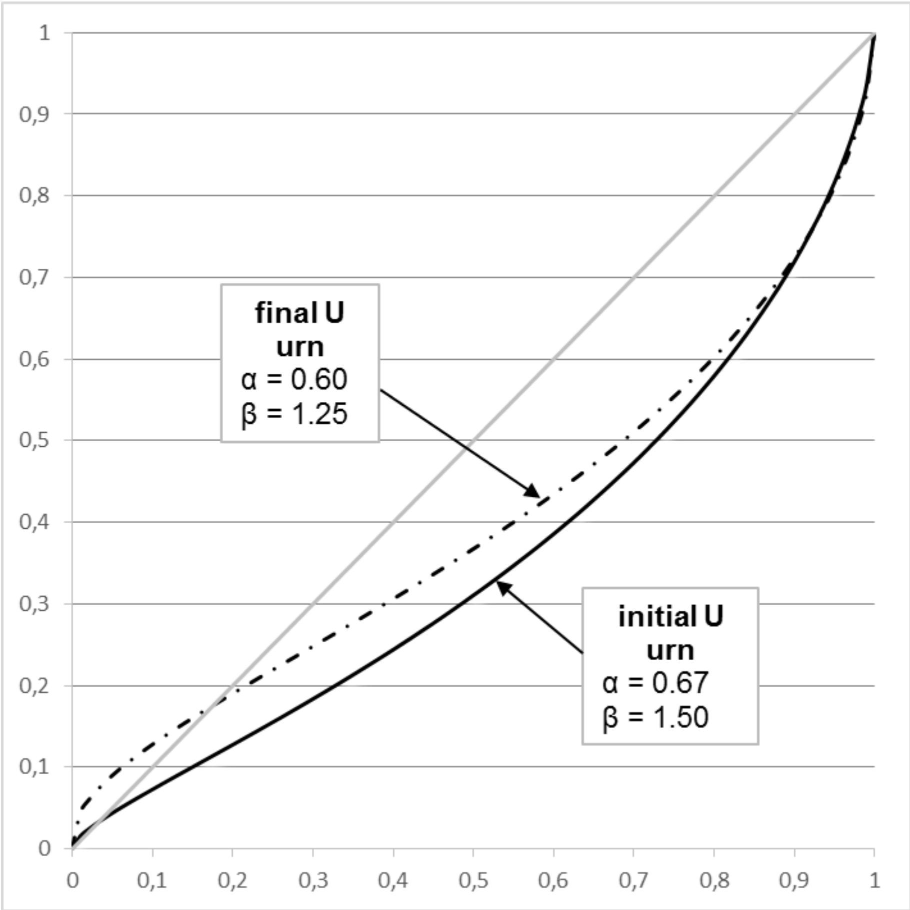

Figure D4. Median individual probability weighting functions: unknown urns, cluster 1

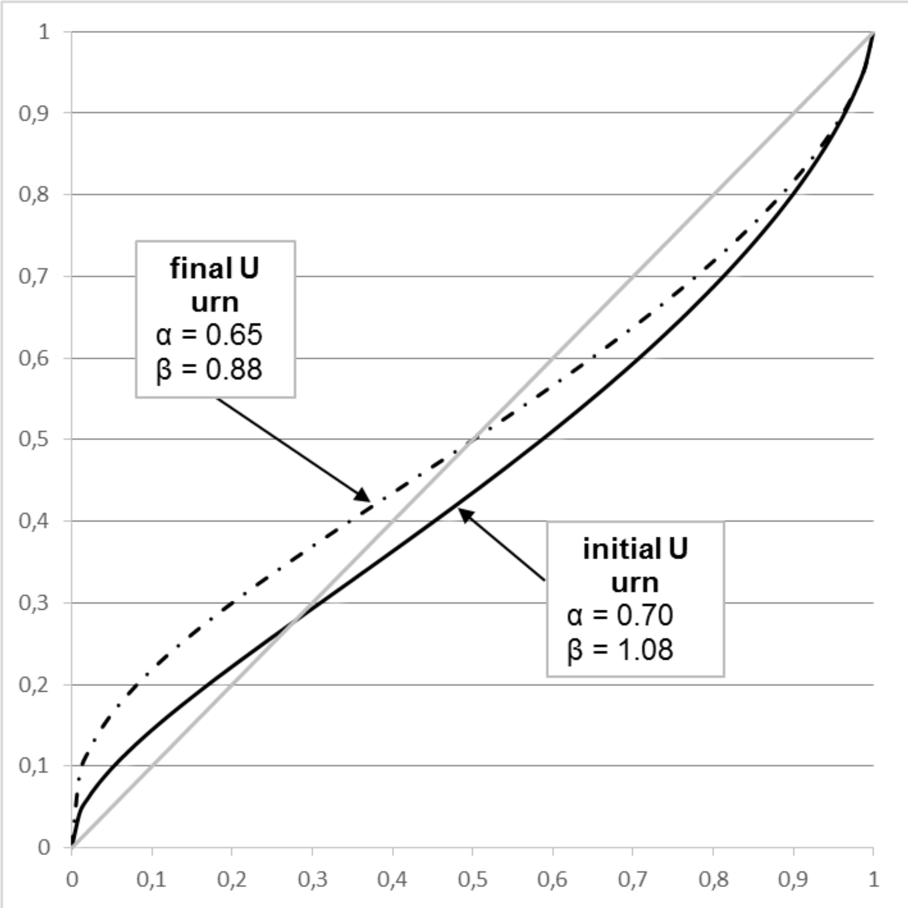

Figure D5. Median individual probability weighting functions: unknown urns, cluster 2

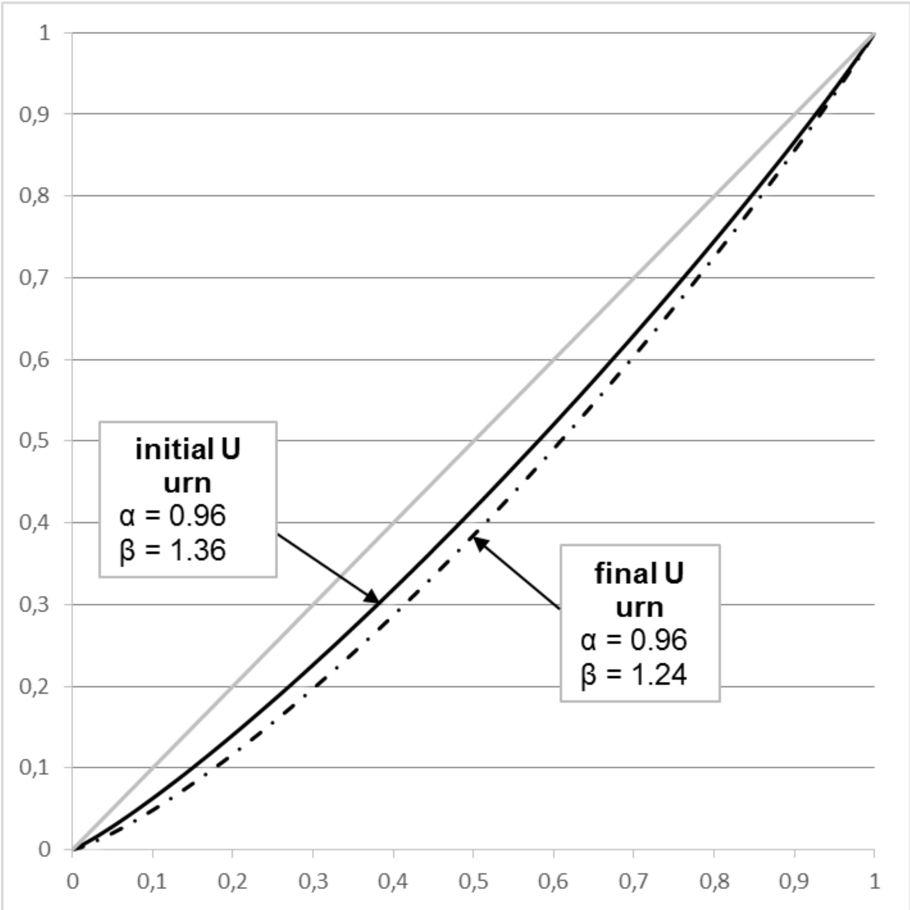

Figure D6. Median individual probability weighting functions: unknown urns, cluster 3.