Categorize Then Choose: Boundedly Rational Choice and Welfare - **Technical and data appendix**

Paola Manzini Marco Mariotti

Queen Mary, University of London and IZA

Marco Mariotti

Queen Mary, University of London

This version: May 2009

A Raw Data

We describe below the variable names used in Table 1:

T: treatment (0 for PAY and 1 for HYP)

SS: session number

SB: subject number

SX: subject's sex (F for Female and M for Male)

Choices between plans are coded as follows: *abcdn* indicates the choice between plans a, b, c and d of length n periods. A value of 1, 2, 3 or 4 indicates that a, b, c or d, respectively, was chosen. Similar for choices *abcn* (involving three plans only) and *abn* (involving two plans only).

T	SS	SB	SX	ki3	id3	dk3	ij3	jk3	jd3	kid3	jki3	djk3	idj3	kidj3	ki2	id2	dk2	ij2	jk2	jd2	kid2	jki2	djk2	idj2	kidj2
0	1	0	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	1	1	F	1	1	2	1	1	1	1	2	3	2	1	1	2	2	2	2	2	1	2	1	3	4
0	1	2	F	2	1	2	1	2	2	2	2	2	1	1	1	1	1	1	2	2	1	3	3	1	1
0	1	3	F	1	1	2	1	2	1	1	3	3	1	1	1	1	2	1	2	2	1	2	3	1	1
0	1	4	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	1	5	F	1	2	1	1	2	2	1	2	3	2	1	1	2	2	1	2	2	1	2	3	2	1
0	1	6	M	1	2	1	1	2	2	3	3	1	2	3	1	2	1	1	1	2	3	2	1	2	3
0	1	7	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	1	8	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	1	9	F	2	1	2	2	1	1	2	1	2	3	4	2	1	2	1	1	1	2	3	2	3	4

0	2	0	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	2	1	F	2	2	1	1	1	2	3	3	1	2	4	1	1	2	1	2	2	1	2	1	2	1
0	2	2	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	2	3	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	2	4	M	1	2	2	1	2	2	1	2	3	2	1	1	2	2	1	2	2	1	2	3	2	1
0	2	5	F	1	1	2	1	2	2	3	2	3	1	1	1	1	2	1	2	2	1	2	3	2	1
0	2	6	M	1	2	1	1	2	2	1	2	1	2	1	1	2	1	1	2	2	3	2	1	2	3
0	2	7	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	2	8	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	2	9	M	2	1	2	1	2	1	2	2	3	1	2	1	1	2	1	2	1	1	3	3	1	2
0	2	10	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	2	11	F	1	2	1	1	2	2	3	2	1	2	3	1	1	1	1	2	2	1	2	1	3	3
0	2	12	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	2	13	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	2	14	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	2	15	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	3	0	M	2	2	2	1	2	2	2	3	1	1	1	2	1	2	1	2	2	2	3	2	1	1
0	3	1	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	3	2	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	3	2	3
0	3	3	F	1	2	2	1	2	2	1	2	3	2	1	1	2	2	1	2	2	1	2	3	2	1

0	3	4	M	1	2	1	1	2	2	1	2	1	2	1	1	2	2	1	2	2	1	2	3	2	1
0	3	5	M	1	2	2	1	2	2	1	2	3	2	1	1	2	1	1	2	2	3	2	1	2	3
0	3	6	F	1	2	2	1	1	2	3	2	3	2	1	1	2	1	1	2	2	1	2	1	2	3
0	3	7	F	1	2	2	1	2	2	1	2	3	2	1	1	2	1	1	2	2	1	2	1	2	3
0	3	8	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	4	0	F	1	2	2	1	2	2	1	2	3	2	1	1	2	1	1	2	2	1	2	1	2	1
0	4	1	M	1	2	1	2	2	2	3	2	1	2	4	1	2	1	1	2	2	3	2	1	2	3
0	4	2	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	4	3	M	1	2	2	1	2	2	1	2	3	2	1	1	2	1	1	2	2	1	2	3	2	1
0	4	4	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	1	2	3	2	1	2	3
0	4	5	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	4	6	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	4	7	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	4	8	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	4	9	M	1	1	2	1	2	1	1	2	3	3	1	1	1	2	1	1	1	1	2	3	1	1
0	4	10	M	1	2	1	1	2	2	3	3	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	4	11	F	1	1	2	1	2	2	1	2	3	1	1	1	1	1	1	2	2	1	3	3	1	3
0	5	0	F	1	1	1	2	2	2	1	3	1	3	2	1	1	1	2	2	2	3	3	1	3	1
0	5	1	F	1	1	2	1	2	1	2	3	2	1	2	2	1	2	2	1	1	1	3	2	1	4
0	5	2	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3

0	5	3	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	5	4	F	1	2	1	1	2	2	3	2	1	2	1	1	2	1	1	2	2	2	2	3	2	1
0	5	5	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	5	6	M	1	2	1	1	2	2	1	2	1	2	1	1	2	1	1	2	2	1	3	1	2	3
0	5	7	F	1	1	2	1	2	1	1	2	3	1	1	2	1	2	1	2	1	2	3	3	1	2
0	6	0	F	1	2	1	1	2	2	1	2	2	2	3	1	2	1	2	2	2	3	2	1	2	3
0	6	1	M	1	2	1	2	1	1	3	1	2	2	4	2	2	1	2	1	2	3	1	1	2	3
0	6	2	F	1	2	1	1	2	2	3	2	1	2	3	1	2	2	1	2	2	3	2	1	2	3
0	6	3	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	6	4	M	1	1	2	1	2	1	1	2	3	1	1	1	1	2	2	2	2	1	2	3	1	1
0	6	5	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	6	6	M	1	1	1	1	1	1	3	2	3	2	1	1	1	2	1	1	2	3	2	1	3	1
0	6	7	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	6	8	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	6	9	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	6	10	F	1	2	1	1	2	2	3	2	1	2	3	1	2	2	1	2	2	3	2	1	2	3
0	6	11	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	3	2	3
0	7	0	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	7	1	F	1	2	2	1	2	2	1	2	3	2	1	1	2	2	1	2	2	1	2	3	2	1
0	7	2	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3

0	7	3	F	1	2	2	1	2	2	1	2	3	2	1	1	2	1	1	2	2	3	2	1	2	1
0	7	4	F	1	2	2	1	2	2	1	2	3	2	1	1	2	2	1	2	2	1	2	1	2	3
0	7	5	M	1	2	2	1	2	2	1	2	3	2	1	1	2	2	1	2	2	1	2	3	2	1
0	7	6	F	1	2	2	1	2	2	1	2	3	2	1	1	2	1	1	2	2	3	2	3	2	1
0	7	7	F	1	2	2	1	2	2	1	2	3	2	1	1	2	2	1	2	2	1	2	3	2	1
0	7	8	F	2	1	2	1	2	1	2	3	3	1	2	2	1	2	1	2	2	2	3	3	1	2
0	7	9	M	1	2	2	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	7	10	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	7	11	M	1	2	1	2	1	2	3	3	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	8	0	M	1	1	2	1	1	1	2	3	1	1	4	2	1	2	1	1	1	2	1	2	3	4
0	8	1	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	8	2	M	1	2	1	2	2	2	3	2	1	2	3	1	2	1	2	1	1	1	1	2	3	4
0	8	3	F	1	2	1	1	2	2	1	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	8	4	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	8	5	M	1	2	2	1	2	1	1	1	1	1	3	1	2	2	2	2	1	3	2	1	2	4
0	8	6	F	1	1	2	1	2	2	1	2	3	1	1	1	1	2	1	2	2	1	2	3	2	1
0	8	7	M	1	2	1	2	2	2	1	2	1	2	3	1	2	1	1	2	2	3	2	2	2	3
0	8	8	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	8	9	M	1	2	2	1	2	2	1	2	3	2	1	1	2	2	1	2	2	3	3	1	2	3
0	8	10	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3

0	9	0	M	1	2	2	2	2	2	1	2	3	3	1	2	2	2	1	2	2	3	2	3	2	1
0	9	1	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	9	2	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	9	3	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	9	4	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
0	9	5	M	1	2	1	1	2	2	3	2	1	2	3	1	2	2	1	1	2	1	2	3	2	1
0	9	6	F	2	1	2	1	1	1	2	3	2	1	2	1	1	2	1	1	2	2	1	3	1	4
0	9	7	M	1	2	1	1	2	2	3	2	1	2	1	1	2	2	1	2	2	3	2	3	2	1
0	9	8	M	1	1	2	1	2	1	1	2	3	1	1	1	1	2	1	2	1	2	2	3	1	1
0	9	9	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	3	2	3
0	9	10	F	1	2	2	1	2	2	1	2	3	2	1	1	2	2	1	2	2	1	2	1	2	1
0	9	11	M	1	1	2	2	2	1	1	2	3	3	1	1	1	2	1	2	1	1	2	3	3	1
1	1	0	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	1	1	F	1	1	2	1	2	2	1	2	3	1	1	1	2	2	1	2	2	1	2	3	1	1
1	1	2	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	1	3	F	1	1	2	1	2	2	1	2	3	1	1	1	1	2	1	2	2	1	2	3	1	1
1	1	4	F	1	2	2	1	2	2	1	2	3	2	1	1	2	2	1	2	2	1	2	3	2	1
1	1	5	F	1	1	2	1	2	2	1	2	3	1	1	1	2	2	1	2	2	1	2	3	2	1
1	1	6	M	2	1	2	1	2	2	2	3	3	1	2	2	1	2	1	2	2	2	3	3	1	2
1	1	7	F	1	2	1	1	2	2	1	2	3	2	3	1	2	2	1	2	2	1	2	3	2	1

1	1	8	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	1	9	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	1	10	F	1	1	1	1	2	2	3	2	3	1	1	1	2	1	1	1	2	1	2	1	1	1
1	1	12	F	1	2	1	1	2	2	3	3	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	1	13	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	1	14	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	1	15	F	1	2	2	1	2	2	3	2	3	2	3	1	2	1	1	2	2	3	2	1	2	3
1	2	0	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	2	1	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	2	2	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	2	3	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	2	4	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	2	5	M	1	1	1	2	2	2	1	1	2	3	1	2	2	2	1	2	1	3	3	3	2	3
1	2	6	M	1	1	2	1	2	1	1	2	3	1	1	1	1	2	1	2	2	1	2	3	1	1
1	2	7	F	1	2	1	1	2	2	1	2	3	2	1	1	2	1	1	2	2	1	2	3	2	1
1	2	8	M	2	1	2	1	1	1	2	3	2	1	2	2	1	2	1	2	1	2	3	3	1	2
1	2	9	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	2	10	F	1	2	2	1	2	2	1	2	3	2	3	1	2	1	1	2	2	3	2	1	2	3
1	3	0	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	3	1	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3

1	3	2	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	1	2	1	2	3
1	3	3	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	3	4	M	2	2	1	1	1	2	3	3	1	2	3	1	1	2	1	2	2	1	2	3	1	1
1	3	5	F	1	1	2	1	2	2	1	2	3	2	3	1	2	2	1	1	2	1	2	3	2	1
1	3	6	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	3	7	F	2	1	2	2	1	1	2	1	2	3	4	2	1	2	2	1	1	2	1	2	3	4
1	3	8	M	1	1	2	1	2	2	1	3	3	2	1	1	2	1	1	2	2	2	3	3	2	3
1	3	9	M	2	1	2	1	1	1	2	3	2	1	2	2	1	2	1	1	1	2	3	2	1	2
1	3	10	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	3	11	F	1	1	2	1	2	2	1	2	3	1	1	1	2	1	1	2	2	1	2	1	2	1
1	3	12	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	3	1	2	2
1	3	13	M	1	2	2	1	2	2	1	3	3	1	3	2	2	1	1	1	2	2	1	2	1	2
1	3	14	F	1	1	2	1	2	1	1	2	3	1	1	1	1	2	1	2	2	1	2	3	1	1
1	3	15	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	4	0	M	2	1	2	2	1	1	2	1	2	3	4	2	1	2	1	1	1	2	3	3	1	2
1	4	1	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	4	2	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	4	3	M	1	1	2	2	2	1	1	2	3	1	1	1	1	2	1	2	2	1	2	3	1	1
1	4	4	F	2	2	2	1	2	1	1	3	3	2	2	2	1	2	1	2	1	2	3	1	3	2
1	4	5	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3

1	4	6	M	1	2	2	1	2	2	1	2	3	2	1	1	2	2	1	2	2	1	2	3	3	1
1	4	7	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	3	2	3
1	4	8	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	1	2	1	2	3
1	4	9	M	1	2	2	1	2	2	1	2	3	2	1	1	2	2	1	2	2	1	2	3	2	1
1	4	10	M	1	2	2	1	2	2	1	2	3	2	1	1	2	2	1	2	2	1	2	1	2	3
1	4	11	F	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3
1	4	12	F	1	2	1	1	2	2	3	2	1	2	1	1	2	1	1	2	2	3	2	1	2	3
1	4	13	M	1	2	1	1	2	2	3	2	1	2	3	1	2	1	1	2	2	3	2	1	2	3

Table 1: Raw data.

B Frequency distribution of choice profiles

Below we present the frequency distribution for the choice functions we have observed with both two and three period sequences (Tables 2 and 3). How should one read these tables? Because it would be impractical to list all the $20,736$ possible choice functions, in Tables 2 and 3 each choice function is coded in the format X-Y, where X refers to binary choice profiles and Y refers to non binary choice profiles. There are 64 possible combinations of choices from binary sets, so X is a number between 1 and 64, with the corresponding choice profiles listed in Table 4. As for choices out of the non-binary sets, there are 324 possible combinations, so Y is a number between 1 and 324, with the corresponding profiles listed in Table 5.

For instance, consider the modal choice profile in both tables 2 and 3, 51-195. From Table 4 one can see that profile 51 corresponds to $KI = 0$, $ID = 1$, $DK = 0$, $IJ = 0$, $JK = 1$ and $JD = 1$, while from Table 5 profile 195 corresponds to $KID = 3$, $JKI = 2$, $DJK = 1$, $IDJ = 2$, $KIDJ = 3$. Thus the corresponding choice function is $\gamma(\{K, I\}) = K$, $\gamma(\{I, D\}) = D$, $\gamma(\{D, K\}) = D$, $\gamma(\{I, J\}) = I$, $\gamma(\{J, K\}) = K$, $\gamma(\{J, D\}) = D$, $\gamma(\{K, I, D\}) = D$, $\gamma(\{J, K, I\}) = K$, $\gamma(\{D, J, K\}) = D$, $\gamma(\{I, D, J\}) = D$ and $\gamma(\{K, I, D, J\}) = D$.

two period sequences	PAY	HYP
5-22	1	0
6-98	0	1
6-107	0	1
6-308	1	0
6-314	1	0
11-307	1	0
14-259	1	0
14-308	0	1
21-23	1	0
21-76	1	0
21-106	1	0

22-107	1	1
22-143	0	1
24-216	0	1
31-276	1	0
35-4	0	1
35-195	2	0
36-92	0	1
37-60	1	0
37-263	1	0
39-49	1	1
44-192	1	0
49-25	1	0
49-187	1	0
49-220	1	0
51-31	1	1
51-33	1	0
51-49	1	1
51-50	1	0
51-51	1	0
51-117	0	1
51-193	2	2
51-195	47	27
51-196	1	0
51-204	1	0
51-213	3	1
51-215	0	1
53-22	1	5
53-31	1	0
53-49	2	0

54-17	1	0
54-107	1	1
55-22	0	1
55-31	1	0
55-49	7	4
55-51	1	0
55-76	0	1
55-193	1	1
55-195	2	0
55-198	1	0
56-51	1	0
57-63	1	0
59-195	1	0
61-22	1	0
63-301	1	0
Total	102	56

Table 2: Frequency distribution of choice functions for two period sequences.

tree period sequences	PAY	HYP
1-51	1	0
5-49	1	0
5-251	1	0
6-98	1	2
11-282	1	0
14-308	1	2
21-22	3	2
21-25	1	0
21-76	1	0

21-98	1	0
22-104	1	0
22-107	1	0
23-163	1	0
24-133	0	1
29-22	0	1
29-76	1	0
36-198	0	1
36-279	1	0
39-51	1	0
43-198	1	0
49-24	0	1
51-31	3	0
51-33	2	1
51-49	1	1
51-193	1	0
51-195	48	27
51-198	2	1
51-202	1	0
51-211	0	1
53-22	2	4
53-24	1	0
53-52	0	1
53-211	0	1
54-14	1	0
54-107	0	1
55-49	14	4
55-187	0	1
55-195	1	0

55-211	0	1
55-213	0	1
56-8	1	0
57-64	0	1
57-142	1	0
59-193	1	0
59-195	1	0
59-276	1	0
63-76	1	0
Total	102	56

Table 3: Frequency distribution of choice functions for three period sequences.

profile of binary choices	KI	ID	DK	IJ	JK	JD
1	0	0	0	0	0	0
2	1	0	0	0	0	0
3	0	1	0	0	0	0
4	1	1	0	0	0	0
5	0	0	1	0	0	0
6	1	0	1	0	0	0
7	0	1	1	0	0	0
8	1	1	1	0	0	0
9	0	0	0	1	0	0
10	1	0	0	1	0	0
11	0	1	0	1	0	0
12	1	1	0	1	0	0
13	0	0	1	1	0	0
14	1	0	1	1	0	0
15	0	1	1	1	0	0

16	1	1	1	1	0	0
17	0	0	0	0	1	0
18	1	0	0	0	1	0
19	0	1	0	0	1	0
20	1	1	0	0	1	0
21	0	0	1	0	1	0
22	1	0	1	0	1	0
23	0	1	1	0	1	0
24	1	1	1	0	1	0
25	0	0	0	1	1	0
26	1	0	0	1	1	0
27	0	1	0	1	1	0
28	1	1	0	1	1	0
29	0	0	1	1	1	0
30	1	0	1	1	1	0
31	0	1	1	1	1	0
32	1	1	1	1	1	0
33	0	0	0	0	0	1
34	1	0	0	0	0	1
35	0	1	0	0	0	1
36	1	1	0	0	0	1
37	0	0	1	0	0	1
38	1	0	1	0	0	1
39	0	1	1	0	0	1
40	1	1	1	0	0	1
41	0	0	0	1	0	1
42	1	0	0	1	0	1
43	0	1	0	1	0	1
44	1	1	0	1	0	1

45	0	0	1	1	0	1
46	1	0	1	1	0	1
47	0	1	1	1	0	1
48	1	1	1	1	0	1
49	0	0	0	0	1	1
50	1	0	0	0	1	1
51	0	1	0	0	1	1
52	1	1	0	0	1	1
53	0	0	1	0	1	1
54	1	0	1	0	1	1
55	0	1	1	0	1	1
56	1	1	1	0	1	1
57	0	0	0	1	1	1
58	1	0	0	1	1	1
59	0	1	0	1	1	1
60	1	1	0	1	1	1
61	0	0	1	1	1	1
62	1	0	1	1	1	1
63	0	1	1	1	1	1
64	1	1	1	1	1	1

Table 4: Possible profiles in binary choice.

profile	KID	JKI	DJK	IDJ	KIDJ
1	1	1	1	1	1
2	2	1	1	1	1
3	3	1	1	1	1
4	1	2	1	1	1
5	2	2	1	1	1
6	3	2	1	1	1

7	1	3	1	1	1
8	2	3	1	1	1
9	3	3	1	1	1
10	1	1	2	1	1
11	2	1	2	1	1
12	3	1	2	1	1
13	1	2	2	1	1
14	2	2	2	1	1
15	3	2	2	1	1
16	1	3	2	1	1
17	2	3	2	1	1
18	3	3	2	1	1
19	1	1	3	1	1
20	2	1	3	1	1
21	3	1	3	1	1
22	1	2	3	1	1
23	2	2	3	1	1
24	3	2	3	1	1
25	1	3	3	1	1
26	2	3	3	1	1
27	3	3	3	1	1
28	1	1	1	2	1
29	2	1	1	2	1
30	3	1	1	2	1
31	1	2	1	2	1
32	2	2	1	2	1
33	3	2	1	2	1
34	1	3	1	2	1
35	2	3	1	2	1

36	3	3	1	2	1
37	1	1	2	2	1
38	2	1	2	2	1
39	3	1	2	2	1
40	1	2	2	2	1
41	2	2	2	2	1
42	3	2	2	2	1
43	1	3	2	2	1
44	2	3	2	2	1
45	3	3	2	2	1
46	1	1	3	2	1
47	2	1	3	2	1
48	3	1	3	2	1
49	1	2	3	2	1
50	2	2	3	2	1
51	3	2	3	2	1
52	1	3	3	2	1
53	2	3	3	2	1
54	3	3	3	2	1
55	1	1	1	3	1
56	2	1	1	3	1
57	3	1	1	3	1
58	1	2	1	3	1
59	2	2	1	3	1
60	3	2	1	3	1
61	1	3	1	3	1
62	2	3	1	3	1
63	3	3	1	3	1
64	1	1	2	3	1

65	2	1	2	3	1
66	3	1	2	3	1
67	1	2	2	3	1
68	2	2	2	3	1
69	3	2	2	3	1
70	1	3	2	3	1
71	2	3	2	3	1
72	3	3	2	3	1
73	1	1	3	3	1
74	2	1	3	3	1
75	3	1	3	3	1
76	1	2	3	3	1
77	2	2	3	3	1
78	3	2	3	3	1
79	1	3	3	3	1
80	2	3	3	3	1
81	3	3	3	3	1
82	1	1	1	1	2
83	2	1	1	1	2
84	3	1	1	1	2
85	1	2	1	1	2
86	2	2	1	1	2
87	3	2	1	1	2
88	1	3	1	1	2
89	2	3	1	1	2
90	3	3	1	1	2
91	1	1	2	1	2
92	2	1	2	1	2
93	3	1	2	1	2

94	1	2	2	1	2
95	2	2	2	1	2
96	3	2	2	1	2
97	1	3	2	1	2
98	2	3	2	1	2
99	3	3	2	1	2
100	1	1	3	1	2
101	2	1	3	1	2
102	3	1	3	1	2
103	1	2	3	1	2
104	2	2	3	1	2
105	3	2	3	1	2
106	1	3	3	1	2
107	2	3	3	1	2
108	3	3	3	1	2
109	1	1	1	2	2
110	2	1	1	2	2
111	3	1	1	2	2
112	1	2	1	2	2
113	2	2	1	2	2
114	3	2	1	2	2
115	1	3	1	2	2
116	2	3	1	2	2
117	3	3	1	2	2
118	1	1	2	2	2
119	2	1	2	2	2
120	3	1	2	2	2
121	1	2	2	2	2
122	2	2	2	2	2

123	3	2	2	2	2
124	1	3	2	2	2
125	2	3	2	2	2
126	3	3	2	2	2
127	1	1	3	2	2
128	2	1	3	2	2
129	3	1	3	2	2
130	1	2	3	2	2
131	2	2	3	2	2
132	3	2	3	2	2
133	1	3	3	2	2
134	2	3	3	2	2
135	3	3	3	2	2
136	1	1	1	3	2
137	2	1	1	3	2
138	3	1	1	3	2
139	1	2	1	3	2
140	2	2	1	3	2
141	3	2	1	3	2
142	1	3	1	3	2
143	2	3	1	3	2
144	3	3	1	3	2
145	1	1	2	3	2
146	2	1	2	3	2
147	3	1	2	3	2
148	1	2	2	3	2
149	2	2	2	3	2
150	3	2	2	3	2
151	1	3	2	3	2

152	2	3	2	3	2
153	3	3	2	3	2
154	1	1	3	3	2
155	2	1	3	3	2
156	3	1	3	3	2
157	1	2	3	3	2
158	2	2	3	3	2
159	3	2	3	3	2
160	1	3	3	3	2
161	2	3	3	3	2
162	3	3	3	3	2
163	1	1	1	1	3
164	2	1	1	1	3
165	3	1	1	1	3
166	1	2	1	1	3
167	2	2	1	1	3
168	3	2	1	1	3
169	1	3	1	1	3
170	2	3	1	1	3
171	3	3	1	1	3
172	1	1	2	1	3
173	2	1	2	1	3
174	3	1	2	1	3
175	1	2	2	1	3
176	2	2	2	1	3
177	3	2	2	1	3
178	1	3	2	1	3
179	2	3	2	1	3
180	3	3	2	1	3

181	1	1	3	1	3
182	2	1	3	1	3
183	3	1	3	1	3
184	1	2	3	1	3
185	2	2	3	1	3
186	3	2	3	1	3
187	1	3	3	1	3
188	2	3	3	1	3
189	3	3	3	1	3
190	1	1	1	2	3
191	2	1	1	2	3
192	3	1	1	2	3
193	1	2	1	2	3
194	2	2	1	2	3
195	3	2	1	2	3
196	1	3	1	2	3
197	2	3	1	2	3
198	3	3	1	2	3
199	1	1	2	2	3
200	2	1	2	2	3
201	3	1	2	2	3
202	1	2	2	2	3
203	2	2	2	2	3
204	3	2	2	2	3
205	1	3	2	2	3
206	2	3	2	2	3
207	3	3	2	2	3
208	1	1	3	2	3
209	2	1	3	2	3

210	3	1	3	2	3
211	1	2	3	2	3
212	2	2	3	2	3
213	3	2	3	2	3
214	1	3	3	2	3
215	2	3	3	2	3
216	3	3	3	2	3
217	1	1	1	3	3
218	2	1	1	3	3
219	3	1	1	3	3
220	1	2	1	3	3
221	2	2	1	3	3
222	3	2	1	3	3
223	1	3	1	3	3
224	2	3	1	3	3
225	3	3	1	3	3
226	1	1	2	3	3
227	2	1	2	3	3
228	3	1	2	3	3
229	1	2	2	3	3
230	2	2	2	3	3
231	3	2	2	3	3
232	1	3	2	3	3
233	2	3	2	3	3
234	3	3	2	3	3
235	1	1	3	3	3
236	2	1	3	3	3
237	3	1	3	3	3
238	1	2	3	3	3

239	2	2	3	3	3
240	3	2	3	3	3
241	1	3	3	3	3
242	2	3	3	3	3
243	3	3	3	3	3
244	1	1	1	1	4
245	2	1	1	1	4
246	3	1	1	1	4
247	1	2	1	1	4
248	2	2	1	1	4
249	3	2	1	1	4
250	1	3	1	1	4
251	2	3	1	1	4
252	3	3	1	1	4
253	1	1	2	1	4
254	2	1	2	1	4
255	3	1	2	1	4
256	1	2	2	1	4
257	2	2	2	1	4
258	3	2	2	1	4
259	1	3	2	1	4
260	2	3	2	1	4
261	3	3	2	1	4
262	1	1	3	1	4
263	2	1	3	1	4
264	3	1	3	1	4
265	1	2	3	1	4
266	2	2	3	1	4
267	3	2	3	1	4

268	1	3	3	1	4
269	2	3	3	1	4
270	3	3	3	1	4
271	1	1	1	2	4
272	2	1	1	2	4
273	3	1	1	2	4
274	1	2	1	2	4
275	2	2	1	2	4
276	3	2	1	2	4
277	1	3	1	2	4
278	2	3	1	2	4
279	3	3	1	2	4
280	1	1	2	2	4
281	2	1	2	2	4
282	3	1	2	2	4
283	1	2	2	2	4
284	2	2	2	2	4
285	3	2	2	2	4
286	1	3	2	2	4
287	2	3	2	2	4
288	3	3	2	2	4
289	1	1	3	2	4
290	2	1	3	2	4
291	3	1	3	2	4
292	1	2	3	2	4
293	2	2	3	2	4
294	3	2	3	2	4
295	1	3	3	2	4
296	2	3	3	2	4

297	3	3	3	2	4
298	1	1	1	3	4
299	2	1	1	3	4
300	3	1	1	3	4
301	1	2	1	3	4
302	2	2	1	3	4
303	3	2	1	3	4
304	1	3	1	3	4
305	2	3	1	3	4
306	3	3	1	3	4
307	1	1	2	3	4
308	2	1	2	3	4
309	3	1	2	3	4
310	1	2	2	3	4
311	2	2	2	3	4
312	3	2	2	3	4
313	1	3	2	3	4
314	2	3	2	3	4
315	3	3	2	3	4
316	1	1	3	3	4
317	2	1	3	3	4
318	3	1	3	3	4
319	1	2	3	3	4
320	2	2	3	3	4
321	3	2	3	3	4
322	1	3	3	3	4
323	2	3	3	3	4
324	3	3	3	3	4

Table 5: Possible profiles in non-binary choice sets.

C Analysis of violations of the axioms, crosstabulated by sequence length

In this section we report in full the crosstabulation of violations of each of the five axioms (Condorcet Consistency, Pairwise Consistency, WARP, Expansion and Weak WARP - we skip R_γ^N acyclicity as it covers exactly the same number of observations as Weak WARP) considered in the main text by sequence length, on which the summary tables in the main text are based. This allows us to measure for each axiom the proportion of subjects failing to satisfy it for either longer or shorter sequences.

In the tables that follow we use a cross (×) to indicate that the axiom is violated and a tick (✓) to indicate that it holds. For each axiom we report: *(i)* within each treatment (i.e. PAY or HYP) whether the proportion of subjects violating the axiom falls with the increase in sequence length in a statistically significant measure, and whether violations for shorter sequences are associated to violations for longer sequences (i.e. 'random mistakes' in the sense of section C.3); and *(ii)* across treatment whether monetary incentives have an effect on the proportions of subjects violating each axiom.

Condorcet Consistency

		PAY				HYP			
		3 periods				3 periods			
		×		✓		×		✓	
		#	%	#	%	#	%	#	%
2 periods	×	18	*17.6*	21	*20.6*	4	*7.1*	10	*17.9*
	✓	12	*11.8*	**51**	*50*	8	*14.3*	**34**	*60.7*

Table 6: Violations of Condorcet Consistency for different sequence length.

- *Within treatment: PAY.* For the PAY treatment, the proportion of subjects violating Condorcet Consistency falls from 38.2% (i.e. 17.6 + 20.6) to 29.4% (i.e. 17.6 + 11.8) as sequence length increases, and this difference is statistically significant at 10%

confidence level (the exact p-value for the McNemar test is 0.081). In addition, we reject the hypothesis of lack of association between violations of Condorcet Consistency in choices over two and three period sequences (the exact mid-p value for the Fisher test is 0.002) - in short, we can reject the hypothesis that violations are due to random mistakes.

- *Within treatment: HYP.* For the HYP treatment, the proportion of subjects violating Condorcet Consistency falls slightly from 25% to 21.4% as sequence length increases, but this difference is not statistically significant (the exact p-value of the McNemar test is 0.407). In addition, we cannot reject the hypothesis of lack of association between violations of Condorcet Consistency in choices over two and three period sequences (the exact mid-p value for the Fisher test is 0.237) - in short, we cannot reject the hypothesis that violations are due to random mistakes.

- *Across treatment: two period sequences.* Comparing across treatments we note that for the two period sequences the percentage of violations of Condorcet Consistency falls when moving from the PAY (38.2%) to the HYP (25%) treatment, and this difference is statistically significant (Fisher test yields an exact mid-p value of 0.047);

- *Across treatment: three period sequences.* Comparing across treatments we note that for the three period sequences the percentage of violations of Condorcet consistency falls when moving from the PAY (29.4%%) to the HYP (21.4%) treatment, and this difference is not statistically significant (Fisher test yields an exact mid-p value of 0.14).

Pairwise Consistency

- *Within treatment: PAY.* For the PAY treatment, the proportion of subjects violating Pairwise Consistency falls from 11.8% to 6.8% as sequence length increases, but this difference is not statistically significant (the exact p-value for the McNemar test is 0.151). In addition, we cannot reject the hypothesis of lack of association between violations of Pairwise Consistency in choices over two and three period sequences

		PAY				HYP			
		3 periods				3 periods			
		×		✓		×		✓	
		#	%	#	%	#	%	#	%
2 periods	×	2	*1.9*	10	*9.9*	2	*3.6*	1	*1.8*
	✓	5	*4.9*	**85**	*83.3*	1	*1.8*	**52**	*92.8*

Table 7: Violations of Pairwise Consistency for different sequence length.

(the exact mid-p value for the Fisher test is 0.113) - in short, we cannot reject the hypothesis that violations are due to random mistakes.

- *Within treatment: HYP.* For the HYP treatment, the proportion of subjects violating Pairwise Consistency stays unchanged at 5.4%. In addition, we reject the hypothesis of lack of association between violations of Pairwise Consistency in choices over two and three period sequences (the exact mid-p value for the Fisher test is 0.003) - in short, we reject the hypothesis that violations are due to random mistakes.

- *Across treatment: two period sequences.* Comparing across treatments we note that for the two period sequences the percentage of violations of Pairwise Consistency falls when moving from the PAY (11.8%) to the HYP (5.4%) treatment, and this difference is statistically significant at 10% confidence level (Fisher test yields an exact mid-p value of 0.100);

- *Across treatment: two period sequences.* Comparing across treatments we note that for the three period sequences the percentage of violations of Pairwise Consistency falls when moving from the PAY (6.9%) to the HYP (5.4%) treatment, and this difference is not statistically significant (Fisher test yields an exact mid-p value of 0.375).

WARP

		PAY				HYP			
		3 periods				3 periods			
		×		✓		×		✓	
		#	%	#	%	#	%	#	%
2 periods	×	19	*18.6*	24	*23.5*	7	*12.5*	9	*16.1*
	✓	11	*10.8*	**48**	*47.1*	6	*10.7*	**34**	*60.7*

Table 8: Violations of WARP for different sequence length.

- *Within treatment: PAY.* For the PAY treatment, the proportion of subjects violating WARP falls from 42.1% to 29.4% as sequence length increases, and this difference is statistically significant (the exact p-value for the McNemar test is 0.020). In addition, we reject the hypothesis of lack of association between violations of WARP in choices over two and three period sequences (the exact mid-p value for the Fisher test is 0.003) - in short, we can reject the hypothesis that violations are due to random mistakes.

- *Within treatment: HYP.* For the HYP treatment, the proportion of subjects violating WARP falls from 28.6% to 23.2% as sequence length increases, but this difference is not statistically significant (the exact p-value of the McNemar test is 0.304). In addition, we reject the hypothesis of lack of association between violations of WARP in choices over two and three period sequences (the exact mid-p value for the Fisher test is 0.017) - in short, we reject the hypothesis that violations are due to random mistakes.

- *Across treatment: two period sequences.* Comparing across treatments we note that for the two period sequences the percentage of violations of WARP falls when moving from the PAY (42.1%) to the HYP (28.6%) treatment, and this difference is statistically significant (Fisher test yields an exact mid-p value of 0.047);

- *Across treatment: three period sequences.* Comparing across treatments we note that for the three period sequences the percentage of violations of WARP falls when moving from the PAY (29.4%) to the HYP (23.2%) treatment, but this difference is not statistically significant (Fisher test yields an exact mid-p value of 0.206).

Expansion

		PAY				HYP			
		3 periods				3 periods			
		×		✓		×		✓	
		#	*%*	#	*%*	#	*%*	#	*%*
2 periods	×	18	*17.6*	21	*20.6*	5	*8.9*	9	*16.1*
	✓	12	*11.8*	**51**	*50*	8	*14.3*	**34**	*60.7*

Table 9: Violations of Expansion for different sequence length.

- *Within treatment: PAY.* For the PAY treatment, the proportion of subjects violating Expansion falls from 38.2% to 29.4% as sequence length increases, and this difference is statistically significant at 10% confidence level (the exact p-value for the McNemar test is 0.081). In addition, we reject the hypothesis of lack of association between violations of Expansion in choices over two and three period sequences (the exact mid-p value for the Fisher test is 0.002) - in short, we can reject the hypothesis that violations are due to random mistakes.

- *Within treatment: HYP.* For the HYP treatment, the proportion of subjects violating Expansion falls from 25% to 23.2% as sequence length increases, but this difference is not statistically significant (the exact p-value of the McNemar test is 0.5). In addition, we cannot reject the hypothesis of lack of association between violations of Expansion in choices over two and three period sequences (the exact mid-p value for the Fisher test is 0.117) - in short, we cannot reject the hypothesis that violations are due to random mistakes.

- *Across treatment: two period sequences.* Comparing across treatments we note that for the two period sequences the percentage of violations of Expansion falls when moving from the PAY (38.2%) to the HYP (25%) treatment, and this difference is statistically significant (Fisher test yields an exact mid-p value of 0.047);

- *Across treatment: three period sequences.* Comparing across treatments we note that for the three period sequences the percentage of violations of Expansion falls when moving from the PAY (29.4%) to the HYP (23.2%) treatment, but this difference is not statistically significant (Fisher test yields an exact mid-p value of 0.206).

Weak WARP

		PAY				HYP			
		3 periods				3 periods			
		×		✓		×		✓	
		#	%	#	%	#	%	#	%
2 periods	×	5	*4.9*	17	*16.7*	3	*5.4*	3	*5.4*
	✓	7	*6.8*	**73**	*71.6*	2	*3.6*	**48**	*85.6*

Table 10: Violations of Weak WARP for different sequence length.

- *Within treatment: PAY.* For the PAY treatment, the proportion of subjects violating Weak WARP falls from 21.6% to 11.7% as sequence length increases, and this difference is statistically significant (the exact p-value for the McNemar test is 0.032). In addition, we can reject at 10% confidence level the hypothesis of lack of association between violations of Weak WARP in choices over two and three period sequences (the exact mid-p value for the Fisher test is 0.051) - in short, we can reject the hypothesis that violations are due to random mistakes.

- *Within treatment: HYP.* For the HYP treatment, the proportion of subjects violating Weak WARP falls from 10.8% to 9% as sequence length increases, but this difference is not statistically significant (the exact p-value of the McNemar test is 0.5). In addition, we reject the hypothesis of lack of association between violations of Weak WARP in choices over two and three period sequences (the exact mid-p value for the Fisher test is 0.003) - in short, we reject the hypothesis that violations are due to random mistakes.

- *Across treatment: two period sequences.* Comparing across treatments we note that for the two period sequences the percentage of violations of Weak WARP falls when moving from the PAY (21.6%) to the HYP (10.8%) treatment, and this difference is statistically significant (Fisher test yields an exact mid-p value of 0.045);

- *Across treatment: three period sequences.* Comparing across treatments we note that for the two period sequences the percentage of violations of Weak WARP falls slightly when moving from the PAY (11.7%) to the HYP (9%) treatment, and this difference is statistically significant (Fisher test yields an exact mid-p value of 0.298).

Based on tables 9-10 we can summarize the overall violations of the axioms considered *by experimental subject*:

	PAY		HYP	
	#	%	#	%
Condorcet Consistency	51	*50*	22	*39.3*
Expansion	51	*50*	22	*39.3*
Weak WARP	29	*28.4*	8	*14.4*
Pairwise Consistency	17	*16.7*	4	*7.2*
WARP	54	*52.9*	22	*39.3*

Table 11: Overall axiom violations.

whereas the outcome of the various tests are summarized in table 18 in the main text.

C.1 Failures of Rational Shortlist Methods by sequence length

We report in tables 12 and 13 the crosstabulation of violations of Expansion and Weak WARP for each choice function in each treatment.

In addition, we also distinguish more finely the number of subjects which violate which axioms in which choice function in tables 14 and 15. In this way we can see that, although

		Rational Shortlist Methods: PAY									
		Expansion				Expansion					
		×		✓		×		✓			
		#	%	#	%	#	%	#	%		
Weak WARP	×	22	*21.6*	0	*0*	12	*11.8*	0	*0*	×	Weak WARP
	✓	17	*16.7*	**63**	*61.7*	18	*17.6*	**72**	*70.6*	✓	
		Two period sequences				Three period sequences					

Table 12: Violations of WARP* and Expansion by sequence length, PAY treatment

		Rational Shortlist Methods: **HYP**									
		Expansion				Expansion					
		×		✓		×		✓			
		#	%	#	%	#	%	#	%		
Weak WARP	×	6	*10.7*	0	0	5	14.3	0	*0*	×	Weak WARP
	✓	8	*14.3*	42	75	8	8.9	**43**	*76.8*	✓	
		Two period sequences				Three period sequences					

Table 13: Violations of WARP* and Expansion by sequence length, HYP treatment

there is no theoretical reason for this to happen, the fact that violations of Weak WARP imply violations of Expansion is an empirical regularity.

Rational Shortlist Methods: PAY

2\3→	both violated		EXP only		Weak WARP only		none	
↓	#	%	#	%	#	%	#	%
both violated	5	*4.9*	8	*7.8*	0	*0*	9	*8.8*
EXP only	3	*2.9*	2	*1.9*	0	*0*	12	*11.8*
Weak WARP only	0	*0*	0	*0*	0	*0*	0	*0*
none violated	4	*3.9*	8	*7.8*	0	*0*	51	*50*

Table 14: Violations of the axioms characterising RSM by sequence length - PAY.

Rational Shortlist Methods: HYP

2\3→	both violated		EXP only		Weak WARP only		none	
↓	#	%	#	%	#	%	#	%
both violated	3	5.4	0	*0*	0	*0*	3	*5.4*
EXP only	0	0	2	*3.6*	0	*0*	6	*10.7*
Weak WARP only	0	0	0	*0*	0	*0*	0	*0*
none violated	2	3.6	6	*10.7*	0	*0*	34	*60.6*

Table 15: Violations of the axioms characterising RSM by sequence length - HYP.

C.2 Failures of Transitive Categorize Then Choose choice functions by sequence length

We report in tables 16 and 17 the crosstabulation of violations of Pairwise Consistency and Weak WARP for each choice function in each treatment.

		Two period sequences					Three period sequences				
		Pairwise Consistency					Pairwise Consistency				
		×		✓			×		✓		
		#	%	#	%		#	%	#	%	
Weak WARP	×	7	*6.9*	15	*14.7*		4	*3.9*	8	*7.8*	×
	✓	5	*4.9*	**75**	*73.5*		3	*2.9*	**87**	*85.3*	✓

TRANSITIVE CATEGORISE THEN CHOOSE: PAY

Table 16: Violations of Weak WARP and Pairwise Consistency by sequence length - PAY

		Two period sequences					Three period sequences				
		Pairwise Consistency					Pairwise Consistency				
		×		✓			×		✓		
		#	%	#	%		#	%	#	%	
Weak WARP	×	0	*0*	6	*10.7*		2	*3.6*	3	*5.4*	×
	✓	3	*5.4*	**47**	*83.9*		1	*1.8*	**50**	*89.3*	✓

TRANSITIVE CATEGORISE THEN CHOOSE: **HYP**

Table 17: Violations of Weak WARP and Pairwise Consistency by sequence length - HYP

C.3 Miscellanea of additional experimental results I: Incentives and sequence length

In this section we report a series of additional considerations that can be drawn from our experimental data, but that are somwhat peripheral to the point we want to make in the paper, and that we excised from it in the interest of space. Nevertheless, we report these here and in the next section for completeness.

Two clear patterns concerning violations of the axioms that emerge from both the aggregate data tables and the individual choice data tables are the following:

1. *People are more consistent for longer sequences.* For *each* axiom considered, the proportion of choices or subjects violating it falls as sequence length increases, irrespective of treatment.

2. *People are more consistent if they are not paid.* For *each* axiom considered, the proportion of choices or subjects violating it falls when incentives are removed, i.e. when passing from the PAY to the HYP treatment, regardless of sequence length.

In addition, by looking at the crosstabulation of violations of each axiom by sequence length, we can measure, for each axiom, the *proportion of subjects* failing to satisfy it for at least one choice function. Crosstabulations of this sort allow us to test for each axiom the following:

- *within each treatment*: *(i)* the statistical significance of the fall in the proportions of violations when going from shorter to longer sequences, and *(ii)* whether or not violations observed for different sequence length are associated.

- *across treatments*: whether, controlling for sequence length, the proportion of violations depends on treatment, i.e. whether elicitation of choices by incentive compatible means in the PAY treatment results in a different proportion of subjects violating each axiom as compared to the HYP treatment.

We summarize our main findings in table 18 below.[1] Notationwise, π_2 and π_3 refer to the proportions of subjects violating an axiom in choices involving two and three period

[1]Ths summary builds on the detailed derivation contained in the previous section.

sequences, respectively, for any given treatment. In addition, for any given sequence length, π_{PAY} and π_{HYP} refer to the proportions of subjects violating an axiom in the PAY and HYP treatment, respectively.

	Within treatment				Across treatment	
	PAY		HYP		$\pi_{PAY} > \pi_{HYP}$	
	$\pi_2 > \pi_3$	random errors	$\pi_2 > \pi_3$	random errors	2 periods	3 periods
CC	✓	×	×	✓	✓	×
PC	×	✓	×	×	✓ (10%)	×
Weak WARP	✓	× (10%)	×	×	✓	×
EXP	✓ (10%)	×	×	✓	✓	×
WARP	✓	×	×	×	✓	×

Table 18: Comparisons of proportions and association.

In the leftmost part of table 18 (under the heading 'within treatment') we report *(i)* whether or not π_2 is statistically larger than π_3, and *(ii)* whether violation of an axiom for shorter sequences makes it any more likely that the subject violates the same axiom when choosing out of longer sequences, too. If this is not the case, one may assume that differences in the proportions of violations across sequence length are due to the subjects making mistakes independently from one another - in table 18 this lack of association is referred to by the shorthand 'random errors'. In each column, we use a tick (✓) when the relevant statistic is such that the heading in the table 'holds', and a cross (×) when the heading in the column 'fails'. So for point *(i)*, a tick indicates that π_2 is statistically larger than π_3, while a cross indicates that the difference in proportions is not statistically significant.[2] Regarding *(ii)* instead we use a tick to indicate that indeed differences may be just random, and a cross when this is not the case.[3]

[2]To be precise, the null hypothesis of the test for *(i)* is that the proportion of violations is the same regardless of sequence length against a one sided alternative that the proportion of violations for two period sequences is larger than for three period sequences. Then the tick refers to the null being rejected. To test this hypothesis we rely on McNemar' statistic.

[3]To be precise, the null hypothesis of the test is for lack of association between rows and columns in the cross-tabulation (i.e. the odds ratio is equal to 1). If this hypothesis is rejected, then rows and

The rightmost part of the table instead reports whether, for any given sequence length, the fall in the proportion of subjects violating a given axiom when moving from the PAY to the HYP treatment is statistically significant (which we denote by a tick ✓) or not (which we denote by a cross ×).[4]

There is no clear pattern of association across sequence length for the violations of each axiom ('random error' columns). Broadly, differences in choice behavior between two and three period sequences are more pronounced in the PAY than in the HYP treatment ('$\pi_2 > \pi_3$' columns). Moreover choice behavior for three period sequences does not differ much between the two treatments, whereas it does for choices over two period sequences ('$\pi_{PAY} > \pi_{HYP}$' columns). However, a quick inspection of choice behavior by subject (table **??**) shows that for all of the three models analyzed, their ability to explain the data increases in the HYP treatment as compared to the PAY treatment: when moving from the PAY to the HYP treatment, the percentage of subjects whose choice function is an RSM increases by 10.7%, the percentage of subjects whose choice function can be rationalized in the standard way increases by 13.6%, and the percentage of subjects whose

columns are associated, i.e. a subject is much more likely to violate the axiom in choice among three period sequences when he has done so in choice out of two period sequences too. In the table we abuse terminology for the sake of clarity, so that a tick corresponds to a *rejection* of the null hypothesis, while a cross stands for failure to reject. We base this test on Fisher's statistic.

Note that tests *(i)* and *(ii)* are independent, in the sense that a high p-value in the McNemar test does not necessarily imply a high p-value of the Fisher test, and viceversa. For instance, in the table

		long	
		yes	no
short	yes	2	8
	no	1	4

based on the Mc Nemar test one rejects the null of equality of proportion whereas based on the Fisher test one fails to reject the null of lack of association between rows and columns.

[4]To be precise, the null hypothesis of the test is for equality in the proportion of subjects violating a given axiom across the PAY and HYP populations, based on the Fisher test (i.e. the odds ratio for the table with treatments against violation is equal to unity). If this hypothesis is rejected (against the one sided alternative that the proportion of violations in the PAY treatment is larger than in the HYP treatment), then the two proportions are statistically different. In the table we abuse notation for the sake of expositional clarity, so that a tick corresponds to a *rejection* of the null hypothesis, while a cross stands for failure to reject.

choice function is CTC increases by 14.1%. For the latter two notions of rationality these increments are statistically significant,[5] and 'just' not significant for rational shortlist methods.[6]

Comment. Our data show a very clear pattern whereby monetary incentives to elicit choices which are the expression of 'true' preferences have the effect of producing less 'rational' behavior. Providing a rigorous explanation for this phenomenon would go beyond the scope of this paper and the bounds of economics. Still, this seems to open a different angle to the discussion on the role of monetary incentives in experiments. In the economics literature this generally revolves around whether or not monetary incentives are necessary to elicit 'true' preferences or the 'best' outcome (see e.g. Camerer and Hogarth [1], Hertwig and Ortmann [3], Read [5] and Harrison and Rutström [2]). However, we note that an empirical regularity in experiments is that subjects are upset when confronted with their own inconsistencies.[7] One might argue that the absence or presence of monetary incentives constitutes a change of (experimental) 'frame', so that what matters is not the composition of the set from which the choice is going to be made, rather the objects it includes *and* whether or not monetary incentives for choice exist. In other words, the choice set when any alternative once chosen is then going to be experienced is a different object from a choice set with the same set of available alternative but where choice itself is just a thought experiment. In addition, there may be other 'external' relevant dimensions to the problem, such as the decision makers' values, motivations and so on, which might influence choice.[8] Based on these considerations, we advance the tentative hypothesis that our results, too, support the position that incentive compatible elicitation of preference is necessary to elicit preferences that are closer to those that a decision maker would display in a real life choice situation. Where choices are only hypothetical in nature, as in the

[5]In comparing proportions of violations in the PAY and HYP treatments, Fisher test's exact mid-p values are 0.042 for WARP, i.e. standard rationalisability; and 0.022 for Weak WARP, i.e. Categorise Then Choose.

[6]The exact mid-p value from the Fisher test is is 0.101.

[7]This is based on the casual evidence that generally emerges in de-briefing discussions, although there is psychological literature that deals with the effect of affective states on decisions, see e.g. Luce, Bettman and Payne [4].

[8]This point has been made very clearly by Sen [6].

HYP treatment, the decision maker's main concern is that of being consistent, resulting in less frequent violations of the axioms. In this sense, our results invite caution in the use of introspection for testing the 'plausibility' of competing axioms of choice.

C.4 Miscellanea of additional experimental results II: 'La donna e' mobile'

When looking at violations of the axioms across sexes, a pattern emerges: within each treatment the proportion of women that violate the axioms is higher than men, albeit with one exception. In addition, the pattern we highlighted in the previous section - whereby with the removal of monetary incentives for choice the proportion of violations of our axioms decreases - persists regardless of sex, as shown in Table 19.

	PAY				HYP			
	F		M		F		M	
	#	%	#	%	#	%	#	%
CC	27	*57.4%*	24	*43.6%*	14	*48.3%*	8	*29.6%*
Expansion	27	*57.4%*	24	*43.6%*	14	*48.3%*	8	*29.6%*
Weak WARP	17	*36.2%*	12	*21.8%*	5	*17.2%*	3	*11.1%*
PC	7	*14.9%*	10	*18.2%*	3	*10.3%*	1	*3.7%*
WARP	27	*57.4%*	27	*49.1%*	14	*48.3%*	8	*29.6%*

Table 19: Violations of the axioms by sex.

The statistical significance of these differences, however, is in general weak. In particular:

1. Within treatment:

    - The difference in proportions of men and women violating Condorcet Consistency and Expansion is statistically significant at 10% confidence level in both the PAY and the HYP treatments;[9]

[9]The mid-p value for the Fisher test is 0.087 for the PAY treatment and 0.084 for the HYP treatment.

- The difference in proportions of men and women violating Pairwise Consistency is not significant in either treatment;[10]
- The difference in proportions of men and women violating Weak WARP is statistically significant at 10% confidence level in the PAY treatment and not statistically significant in the HYP treatment;[11]
- The difference in proportions of men and women violating WARP is not statistically significant in the PAY treatment and statistically significant at 10% confidence level in the HYP treatment.[12]

2. across treatments:

   - For female participants, the only difference in the proportions of subjects violating a given axiom across treatments which is statistically significant is for Weak WARP, for which the Fisher test returns a mid-p value of 0.042;[13]
   - For male participants the only differences in proportions which are statistically significant are for Pairwise Consistency, for which the Fisher test yields a p-value of 0.037; and WARP, for which the mid-p value from the Fisher test is 0.051.[14]

The analysis above shows also that there are substantial differences (though not always statistically significant) in the proportion of women and men whose choices conform to either Full rationality (i.e. their choices satisfy WARP) or RSM by categorization (i.e. their choices satisfy Weak WARP). To check differences in the sexes as to the explanatory power of RSM we present cross-tabulations of Expansion and Weak WARP by sex in tables 20 and 21.

[10]The mid-p value for the Fisher test is 0.336 for the PAY treatment and 0.199 for the HYP treatment.

[11]The mid-p value for the Fisher test is 0.059 for the PAY treatment and 0.27 for the HYP treatment.

[12]The mid-p value for the Fisher test is 0.21 for the PAY treatment and 0.084 for the HYP treatment.

[13]The mid-p value for the Fisher test is equal to 0.224 for Condorcet Consistency, WARP and Expansion, and equal to 0.30 for Pairwise Consistency.

[14]The other mid-p values for the Fisher test are equal to 0.117 for both Condorcet Consistency and Expansion, and to 0.129 for Weak WARP.

PAY

		Females: Expansion × #	× %	✓ #	✓ %	Males: Expansion × #	× %	✓ #	✓ %		
Weak WARP	×	17	36.2%	0	0%	12	21.8%	0	0%	×	Weak WARP
	✓	10	21.3%	20	42.6%	12	21.8%	31	56.4%	✓	

Table 20: RSM by sex in the PAY treatment.

HYP

		Females: Expansion × #	× %	✓ #	✓ %	Males: Expansion × #	× %	✓ #	✓ %		
Weak WARP	×	5	17.2%	0	0%	3	11.1%	0	0%	×	Weak WARP
	✓	9	31.1%	15	51.7%	5	18.5%	19	70.4%	✓	

Table 21: RSM by sex in the HYP treatment.

When considering RSMs, then, the difference across sexes is quite substantial in both treatments (around 13% in the PAY treatment and just short of 20% in the HYP treatment), and it is also statistically significant at 10% confidence level for both treatments.[15] Finally, differences across treatments by sex are not statistically significant.[16]

Turning next to the Transitive CTC model distinguishing by sex:

PAY

Females — Pairwise Consistency

		×		✓	
		#	%	#	%
Weak WARP	×	6	12.82%	11	23.4%
	✓	1	2.1%	29	61.7%

Males — Pairwise Consistency

×		✓			
#	%	#	%		
5	9.1%	7	12.7%	×	Weak WARP
5	9.1%	38	69.1%	✓	

Table 22: Transitive CTC by sex in the PAY treatment

HYP

Females — Pairwise Consistency

		×		✓	
		#	%	#	%
Weak WARP	×	1	3.4%	4	13.8%
	✓	2	6.9%	22	75.9%

Males — Pairwise Consistency

×		✓			
#	%	#	%		
1	3.7%	2	7.4%	×	Weak WARP
0	0%	24	88.9%	✓	

Table 23: Transitive CTC by sex in the HYP treatment

In summary then:

[15]The mid-p values for the Fisher test of the difference in the proportion of men and women satisfying RSM is equal to 0.087 for the PAY treatment and 0.084 for the HYP treatment.

[16]The mid-p values for the Fisher test of the difference in the proportion of subjects satisfying RSM in the HYP and PAY treatments is equal to 0.228 for Female participants and equal to 0.117 for Male participants.

	PAY				HYP			
	F		M		F		M	
	#	%	#	%	#	%	#	%
Full rationality	20	*42.6*	28	*50.9*	15	*51.7*	19	*70.4*
Rational Shortlist Methods	20	*42.6*	31	*56.4*	15	*51.7*	19	*70.4*
Categorise Then Choose/Rationalisation	30	*63.8*	43	*78.2*	24	*82.8*	24	*88.9*
Order Rationalisation	30	*63.8*	43	*78.2*	24	*82.8*	24	*88.9*
Transitive CTC	29	*61.7*	38	*69.1*	22	*75.9*	24	*88.9*

Table 24: Explanatory power of competing theories across sexes

The last three theories work better than the other two regardless of sex. Across sexes, there are differences:

- the proportion of men whose choices are fully transitive is higher than the proportion of women in the HYP treatment, but not in the PAY treatment;[17]
- the proportion of men whose choices are RSM is statistically significantly higher than the proportion of women in both the PAY and HYP treatment;[18]
- the proportion of men whose choices are an CTC rational is higher than the proportion of women in the PAY but not in the HYP treatment.[19]
- the proportion of men whose choices are an CTC rational is not statistically different than the proportion of women in both treatments.[20]

It is very hard to explain any of these differences in behavior within a purely economic framework. We leave further analysis to scholars in other fields.

[17]The Fisher test yields an exact mid-p value equal to 0.215 for the PAY treatment and to 0.084 for the HYP treatment (i.e. for the latter the difference in proportions is significant at 10% confidence level).

[18]The Fisher test yields an exact mid-p value equal to 0.086 for the PAY treatment and to 0.084 for the HYP treatment (i.e. the difference in proportions is significant at 10% confidence level).

[19]The Fisher test yields an exact mid-p value equal to 0.059 for the PAY treatment (i.e. statistical significance is at 10% confidence level) and to 0.272 for the HYP treatment.

[20]The Fisher test yields an exact mid-p value equal to 0.222 for the PAY treatment and to 0.114 for the HYP treatment.

References

[1] Camerer, C. F. and R. Hogarth (1999) "The effects of financial incentives in experiments: A review and capital-labor-production framework", *Journal of Risk and Uncertainty*, 19: 7-42.

[2] Harrison, G. and E. Rutström "Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods", in C. Plott and V. Smith, eds., *Handbook of Experimental Economics Results*. New York: Elsevier, forthcoming.

[3] Hertwig, R. and A. Ortmann (2001) 'Experimental Practices in Economics: A Challenge for Psychologists?' [target article], *Behavioral and Brain Sciences*, 24: 383-403.

[4] Luce, M., J. Bettman, and J. Payne (1997) "Choice Processing in Emotionally Difficult Decisions", *Journal of Experimental Psychology: Learning, Memory and Cognition,* 23:384-405.

[5] Read, D. (2005) "Monetary incentives, what are they good for?", *Journal of Economic Methodology*, 12: 265-276.

[6] Sen, A. (1993) "Internal consistency of choice", *Econometrica*, 61: 495-521.