Web Appendix to:

An Experimental Test of Prospect Theory for Predicting Choice under Ambiguity

Amit Kothiyal[a], Vitalie Spinu[b], & Peter P. Wakker[c]

[a] Max Planck Institute for Human Development,

Lentzeallee 94, 14195 Berlin, Germany

[b] UCLA Anderson School of Management,

Cornell Hall D526, 110 Westwood Plaza, Los Angeles, CA 90095

[c] Econometric Institute, Erasmus University,

P.O. Box 1738, Rotterdam, 3000 DR, the Netherlands; wakker@ese.eur.nl

Journal of Risk and Uncertainty, in press

November 22, 2013

We discuss the fitting procedure in more detail, and report the median values of individual parameters that we estimated for various models. For source prospect theory (SPT) we provide all the parameter estimates for all subjects. We conclude with the model tournament table computed based on different fitting/test splits of the data (cross validation) than that of HLM (Hey, Lotito, & Maffioletti 2010). References and notation are as in the main text.

1 Additional remarks on the estimation method

As pointed out in the main text, all models were fit individual by individual, and predicted likelihoods were used to compare the models. We here follow the technique used by HLM. Whenever they had done calculations, we used their results, and did not recalculate. Some additional remarks are due.

The advantage of fitting the data individual by individual is mainly computational – the number of parameters estimated are in the range of 3-10 per subject, which given the amount of choice data (135 binary responses per subject) gives a stable fit for most cases. Moreover, the maximum likelihood optimization algorithm (we used Nelder-Mead) converges fast. The disadvantage of individual fitting is two-fold.

First, individual fitting does not take into account the information about the choices and parameters of other subjects. In statistics, taking such information into account is known as a collective inference – given the parameters or choices of a group of subjects we can infer the parameters and predict choices of other individuals. Thus, it may be interesting to to pool all the data and estimate the parameters for all individuals at the same time. This procedure obviously requires additional assumptions at the population level. Mixed effect models and hierarchical Bayesian inference are two standard approaches that can be used. We estimated a mixed effect for SPT and it gave results virtually identical to the individual by individual estimates. Because the optimization procedure is very complex (240 parameters and more than 400 inequality restrictions) and does not give an obvious gain over the individual by individual estimation we did not pursue this method further.

The second problem of individual by individual estimation is that it can lead to overfitting. It does so for several subjects for almost all models we considered. It happens when a subject's choice data (used for fitting) is fit "too" well by a deterministic model and, hence, the σ parameter (representing the model fitting error) becomes too small. As a consequence the model generalizes poorly and predicts very poorly on the test set. To address this issue, HLM removed subject 35 who gave very bad predicted log-likelihoods for multiple prior models.[1] An alternative way out is to use more robust measures of central tendency, such as medians or

[1] Tables WA1 and WA2 below are counterparts to Tables 1 and 2 from the main text with the subject 35 included.

trimmed means, and to perform non-parametric statistical tests without removing any subjects, as reported in the main text.

To reduce the impact of overfitting, HLM also restricted the error parameter σ by imposing the lower bound of 0.01 for all models. We follow the same strategy in our estimations. This is not a serious restriction as it affects only a small number of individuals for our main model, SPT, as can be seen in Figure WA1. It is also empirically plausible to assume a minimal level of error and no perfect fit. In general, the more parameters a model has, the more prone it is to overfitting, and the more subjects will have an estimated sigma equal to the lower bound. We similarly imposed an additional restriction on prospect theory models – we did not allow the individual loss aversion parameter λ to exceed 40. Again, this is empirically plausible and it avoids degenerate estimates.

We also reproduce Tables 1 and 2 from the main text with subject 35 included. The MnEU and αMM models are seriously affected by this inclusion.

	SPT	MxEU	EU	DFT	CEU	EV	MnEU	αMM	MaxMin	MaxMax	MinReg
All	-3.94^b	-4.02	-4.52	-4.70	-6.15	-11.49	-12.57	-12.58	-13.32	-13.97	-14.55
Tr 1	-2.60^b	-2.82	-3.44	-3.50	-3.22	-12.15	-3.13	-2.79	-12.57	-14.52	-14.48
Tr 2	-5.16	-5.14^b	-5.61	-5.93	-6.35	-11.03	-5.70	-5.58	-14.17	-14.45	-14.86
Tr 3	-3.90^b	-3.96	-4.39	-4.53	-8.68	-11.34	-28.73	-29.19	-13.14	-12.94	-14.29

Table WA1: Mean predicted log-likelihoods for the three treatments, and overall (subject 35 included). The biggest (least negative), indicated by superscript b, is the best in each row.

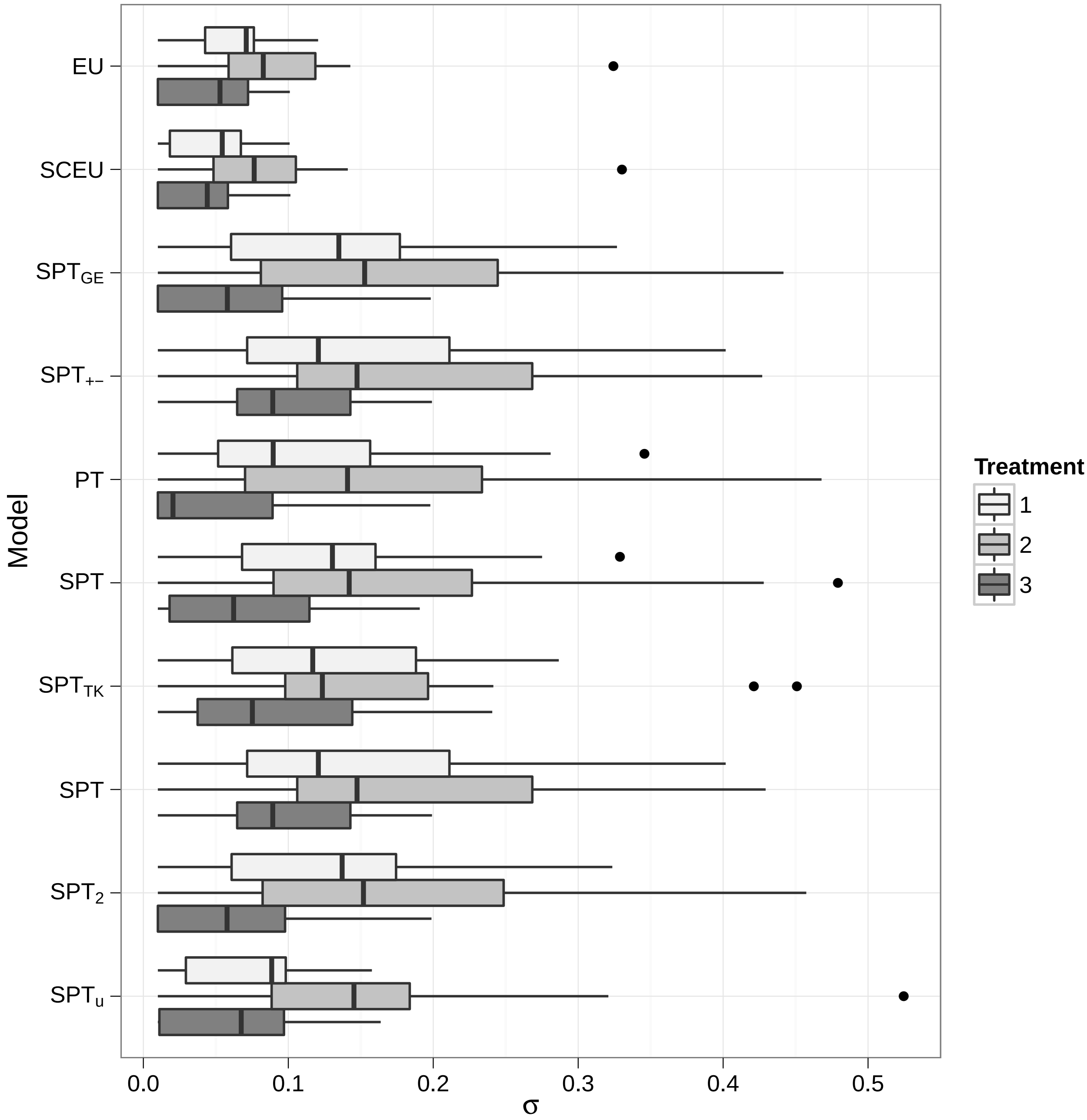

Figure WA1: Boxplots of estimats of the error parameter σ for all variations of PT considered in the paper. The body of each box is formed by 25%, 50% and 75% quantiles. The upper whisker is given by the largest observation smaller than 25% quantile plus 1.5*IQR (inter-quantile range). Observations larger than this number are considered outliers. Lower whiskers are defined symmetrically.

	SPT	DFT	CEU	MxEU	αMM	EU	MnEU	MaxMax	EV	MaxMin	MinReg
All	10^b	8	7	7	6	5	4	1	0	0	0
Tr 1	2	3	3	4^b	1	0	2	0	0	0	0
Tr 2	5^b	2	3	1	2	4	0	0	0	0	0
Tr 3	3^b	3^b	1	2	3^b	1	2	1	0	0	0

Table WA2: Number of subjects for whom a theory predicts best (subject 35 included). The biggest number, indicated by superscript b, is the best in each row.

2 Predicted likelihoods for Prospect Theory models

Table WA3[2] presents medians and trimmed means of predicted likelihood for all variations of prospect theory.

3 Parameter estimates

This section reports on the actual parameter values that we obtained when fitting PT. Table WA4 reports the median parameters estimated based on the fit/test data split of HLM. The meaning of the parameters is as follows. Subjective probabilities of the source method are denoted p_1, p_2, non-additive weights for general GPT are denoted $w_1, \ldots, w_{23}$, the utility parameter is u (normalized utility of 10, where $U(100) = 1$), loss aversion is λ (normalized utility of -10), α and β are the model dependent parameters of the weighting functions as defined in the main text, α_- is the parameter of the Prelec one-parameter weighting used for negative outcomes in $\text{SPT}_\pm$, σ is the estimated value of the error parameter, and $lfit$ and *lpred* are the medians of the fitted and predicted likelihoods of the model concerned.

It is remarkable that the utility parameter (loss aversion in PT) is very large for all the models. In models that do not allow for loss aversion, utility tries to compensate for loss

[2]The aggregated predicted likelihoods of SPT and $SPT_\pm$ differ in their third decimals. Individual predicted likelihoods are also very close.

	mean$_{.1}$	mean$_{.05}$	mean	median
SPT	-3.53^b	-3.62^b	-3.83^b	-3.32
SPT$_\pm$	-3.53^b	-3.62^b	-3.83^b	-3.32
SPT$_u$	-3.56	-3.67	-3.92	-3.37
SCEU	-3.63	-3.73	-3.98	-3.37
SPT$_{NA}$	-3.64	-3.91	-6.21	-3.13
SPT$_2$	-3.87	-4.06	-4.91	-3.03^b
SPT$_{GE}$	-3.92	-4.09	-4.28	-3.12
SPT$_{TK}$	-4.19	-4.27	-4.39	-4.16
EU	-4.34	-4.36	-4.43	-4.33
GPT	-5.10	-5.67	-14.42	-3.84
SCEV	-11.69	-11.61	-11.41	-11.61
SPT$_{\lambda=1}$	-11.73	-11.66	-11.45	-12.04

Table WA3: Means, trimmed means and medians for all PT models (sorted on trimmed mean$_{.1}$). The biggest (least negative), indicated by superscript b, is the best in each column.

aversion. We inspected the individual choices, and they revealed that subjects are indeed extremely loss averse. They mostly minimized the likelihood of losing, almost without trading it off against gaining £100 instead of £10. For example, in the choice between (Yellow: 100, Blue: -10, Pink: -10) and (Yellow: 10, Blue: 10, Pink: -10), the numbers of subjects who preferred the former and the latter were, respectively: 2 versus 13 in Treatment 1, 2 versus 15 in Treatment 2, and 4 versus 12 in Treatment 3. The majority preferences weight a loss of -10 instead of the middle outcome 10 on the unlikely event blue way more than a gain of 100 versus the middle outcome 10 on the likely event yellow. For instance, if probability weighting plays no role and if utility is linear outside of 0, then the majority preferences in this choice imply loss aversion to exceed $(9 - 1) \times 5/3) > 13$. Thus, the very high loss aversion found reflects a genuine phenomenon in the data and is not a misestimation due to the model formulation or to the fitting procedure.

	p_1	p_2	u	λ	α	β	w_1	w_2	w_3	w_{12}	w_{13}	w_{23}	α_-	σ	lfit	lpred
SPT_2	0.23	0.44		1.38	1.29	1.07								0.10	18.78	-3.15
SPT_{GE}	0.23	0.44		1.25	1.29	0.96								0.10	18.74	-3.16
SPT_{NA}	0.23	0.43		1.30	-0.09	-0.07								0.11	19.06	-3.20
SPT	0.23	0.44		1.52	1.22									0.12	19.96	-3.35
$SPT_{\pm}$	0.23	0.44		1.52	0.88								1.22	0.12	19.96	-3.35
SPT_u	0.23	0.44	0.28	0.94	1.27									0.09	19.96	-3.38
SCEU	0.23	0.44	0.61		1.27									0.05	19.96	-3.38
SPT_{TK}	0.23	0.44		1.12	1.32									0.11	20.38	-4.16
GPT				1.42			0.18	0.39	0.31	0.73	0.61	0.84		0.09	16.36	-4.35
EU	0.22	0.44	0.56											0.07	23.15	-4.39
SCEV	0.18	0.49			1.01									0.38	67.50	-11.73
$SPT_{\lambda=1}$	0.17	0.49			0.97									0.38	67.34	-12.11

Table WA4: Medians of estimated parameters across subjects

For our central model, SPT, the estimated parameters are provided in Table WA5. Recall that treatment 1 is associated with the least ambiguity, and treatment 3 with the most. This is confirmed by the variation of the estimated belief parameters, which increases with the ambiguity of the treatment. Boxplots in Figures WA2 and WA3 illustrate this point for subjective probabilities of the events $\{pink\}$ and $\{blue\}$. There is less agreement on the probabilities of the events in the third treatment than in the first two.

	σ	p_1	p_2	λ	α	lfit	lpred
1	0.01	0.24	0.48	4.00	0.31	5.94	-0.00
2	0.10	0.23	0.41	0.52	1.36	25.85	-2.42
3	0.10	0.23	0.51	1.08	1.46	18.94	-2.77
4	0.06	0.22	0.45	0.33	0.80	17.65	-3.38
5	0.12	0.21	0.45	4.00	1.03	7.74	-0.12
6	0.01	0.32	0.34	1.22	0.71	10.64	-1.68
7	0.01	0.33	0.33	0.79	1.04	24.66	-6.04
8	0.13	0.23	0.48	1.74	1.54	19.48	-2.55
9	0.23	0.21	0.48	3.41	1.63	17.96	-2.09
10	0.19	0.19	0.41	1.57	1.17	28.32	-4.54
11	0.14	0.25	0.43	1.74	1.58	22.54	-1.47
12	0.40	0.23	0.44	3.38	1.66	32.52	-2.72
13	0.25	0.23	0.45	4.00	1.06	14.57	-0.95
14	0.09	0.21	0.44	0.53	1.46	20.00	-3.37
15	0.28	0.26	0.41	3.28	1.02	25.72	-4.92
16	0.43	0.50	0.50	0.45	1.06	59.56	-9.57
17	0.07	0.23	0.43	1.94	1.29	9.46	-2.60
18	0.08	0.23	0.46	1.03	1.42	17.32	-3.58
19	0.11	0.21	0.45	1.86	1.16	15.26	-4.30
20	0.15	0.24	0.42	1.94	1.24	20.12	-4.76
21	0.28	0.22	0.43	2.14	1.80	33.17	-4.29
22	0.27	0.25	0.44	1.09	1.40	45.74	-5.83
23	0.13	0.28	0.36	1.59	1.28	20.48	-5.77
24	0.21	0.22	0.42	0.99	2.27	34.43	-12.40

25	0.01	0.25	0.42	0.10	1.32	0.21	-0.05
26	0.11	0.22	0.40	1.51	1.10	19.07	-3.41
27	0.30	0.21	0.46	2.48	1.26	30.16	-3.52
28	0.38	0.22	0.44	4.00	1.29	22.29	-12.08
29	0.14	0.18	0.48	0.48	2.01	26.24	-2.51
30	0.18	0.23	0.42	0.46	0.90	45.63	-8.02
31	0.19	0.19	0.45	1.66	1.43	25.06	-2.17
32	0.10	0.24	0.41	0.90	1.50	22.17	-2.91
33	0.11	0.17	0.46	1.18	0.95	19.24	-3.04
34	0.16	0.23	0.41	2.10	0.73	22.60	-7.94
35	0.06	0.30	0.37	0.67	1.08	20.10	-9.07
36	0.09	0.32	0.34	1.00	1.15	27.47	-8.39
37	0.16	0.16	0.48	2.80	0.52	18.33	-2.16
38	0.19	0.08	0.54	1.53	1.09	21.34	-4.64
39	0.01	0.33	0.34	0.02	1.17	15.40	-6.63
40	0.08	0.30	0.37	4.00	1.21	12.85	-0.31
41	0.01	0.30	0.35	0.29	0.68	5.36	-1.17
42	0.20	0.21	0.47	4.00	1.73	15.34	-1.07
43	0.01	0.33	0.34	0.84	1.26	11.93	-2.40
44	0.07	0.17	0.46	0.16	0.88	19.91	-5.66
45	0.09	0.20	0.45	4.00	0.99	6.12	-0.01
46	0.14	0.20	0.44	0.49	1.37	32.10	-2.70
47	0.09	0.27	0.40	1.55	1.32	16.50	-3.89
48	0.11	0.11	0.54	0.81	1.00	19.73	-3.32

Table WA5: Individual parameters for SPT for all 48 subjects

It is also remarkable that the probability weighting parameters (reflecting ambiguity attitudes here) in general suggest more S-shaped weighting than inverse S-shaped weighting, deviating from the common findings in the literature (also for ambiguity). This is similar to HLM's finding that maxmax EU fits very well, and maxmin does worse. Both these findings suggest that these data contain more optimism than pessimism in event weighting, in deviation from findings in other papers. HLM give no explanation for this unusual finding. We have no explanation for it either, and we can only confirm HLM's finding here.

4 Cross Validation

As an additional test of our main hypothesis (that SPT performs best on this given dataset) and to rule out the posibility that our findings are a consequence of the particular fit/test data split, we performed a cross-validation analysis. We randomly split the total of 162 questions in 10 roughly equal batches. For each batch $i \in \{1, \ldots, 10\}$ we estimated the parameters of all the models excluding batch i. Then we computed the predicted log-likelihood on the test batch i. To aggregate the predicted log-likelihood across 10 batches, we used medians (Table WA6) and trimmed(0.1) means (Table WA7). We also incorporated GPT_u, which adds a utility parameter to GPT. The results of this analysis are very similar to those presented in the main text. This finding confirms that the choice of the test and prediction samples of HLM, followed by us in the main text, are representative.

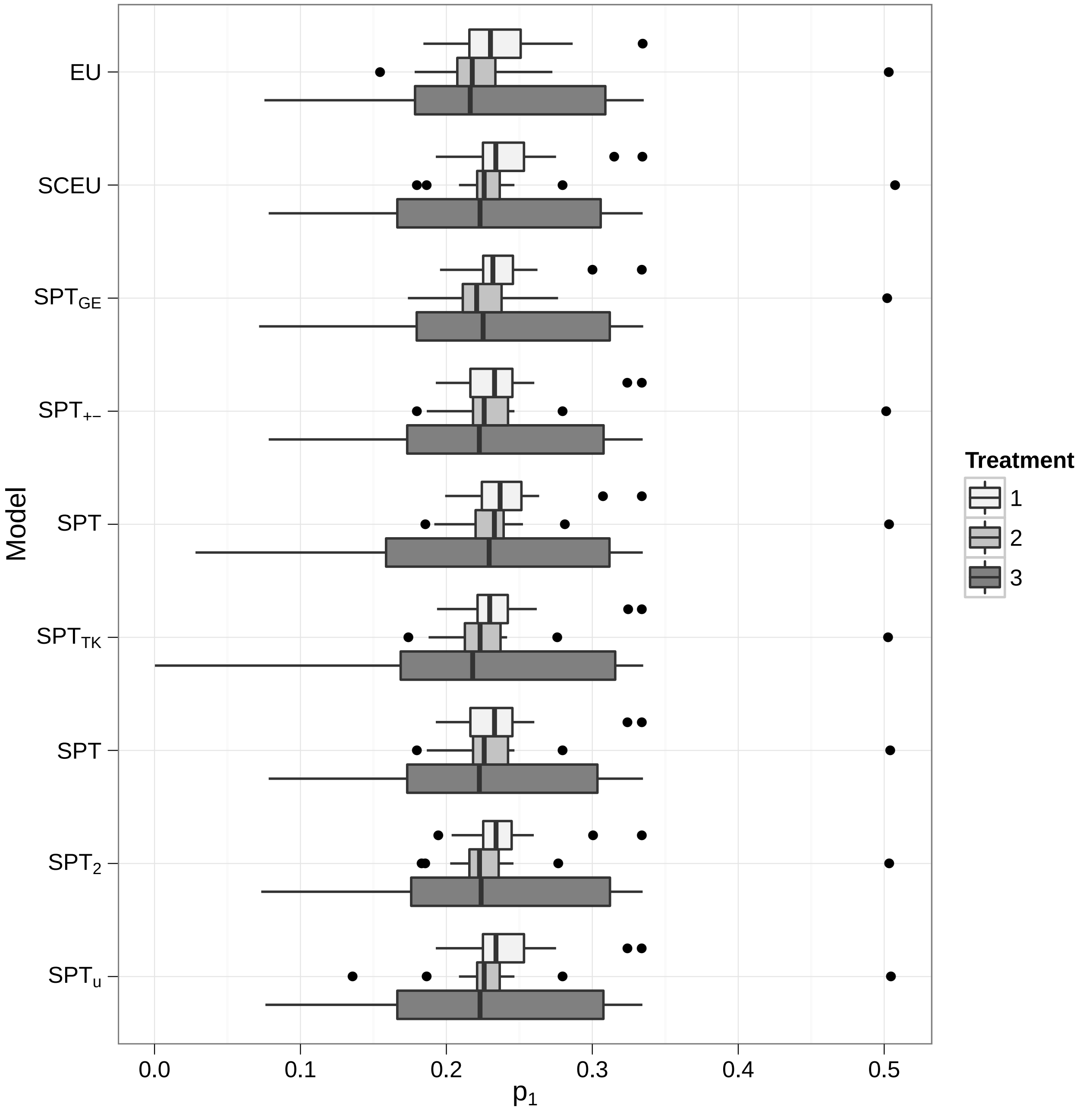

Figure WA2: Individual estimates of the subjective probability of the pink ball (objective probability 0.2)

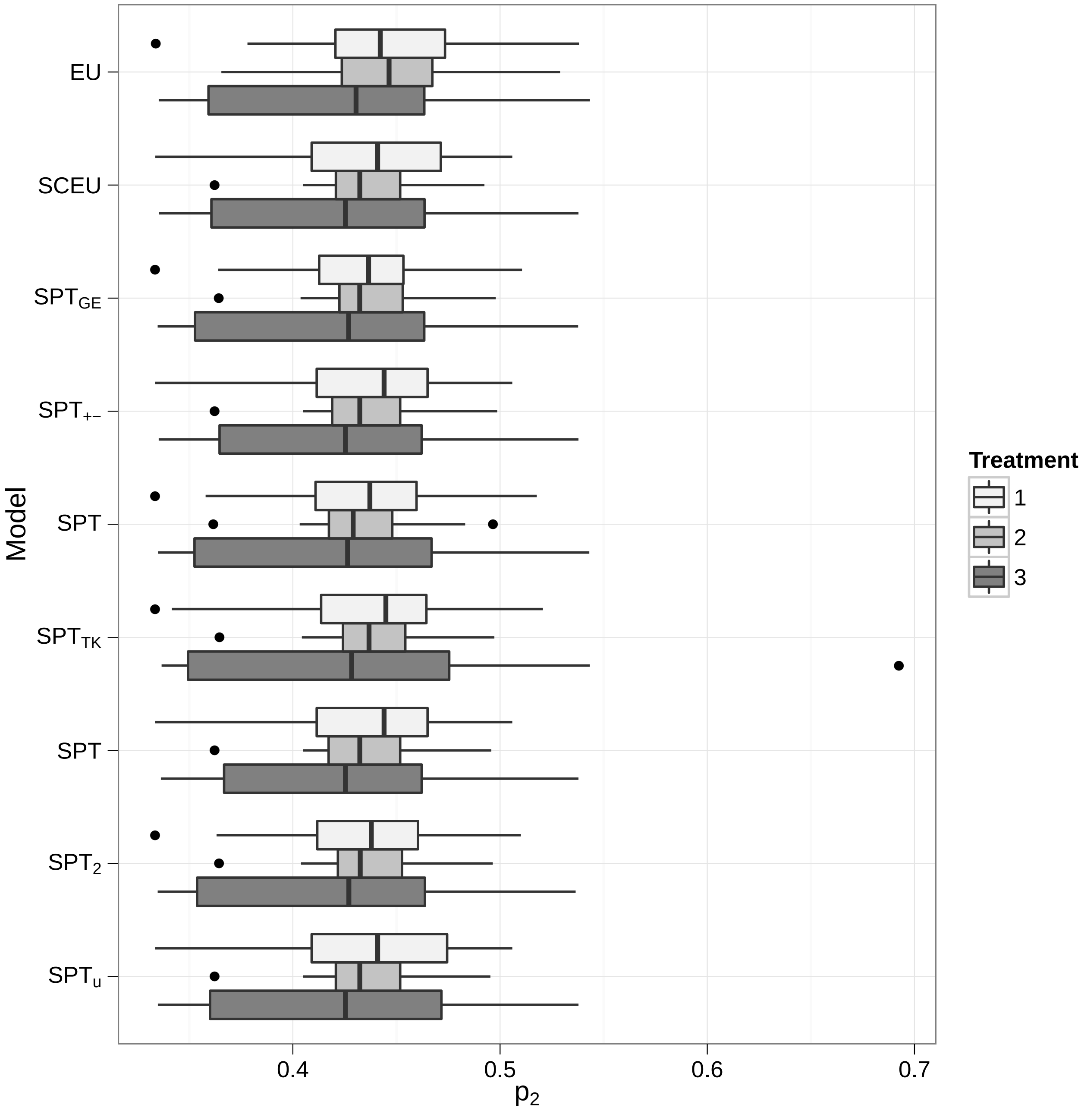

Figure WA3: Individual estimates of the subjective probability of the blue ball (objective probability 0.3)

	SPT_u	SPT	SCEU	GPT_u	CEU	GPT	EU	MnEU	MxEU	αMM	CEV	SCEV
SPT	26^{22}	-	-	-	-	-	-	-	-	-	-	-
SCEU	24^{24}	24^{24}	-	-	-	-	-	-	-	-	-	-
GPT_u	18^{30}	20^{28}	18^{30}	-	-	-	-	-	-	-	-	-
CEU	16^{32}	19^{29}	17^{31}	23^{25}	-	-	-	-	-	-	-	-
GPT	16^{32}	18^{30}	17^{31}	22^{26}	${19^{*}}^{29}$	-	-	-	-	-	-	-
EU	${19^{*}}^{29}$	${20^{*}}^{28}$	20^{28}	21^{27}	21^{27}	22^{26}	-	-	-	-	-	-
MnEU	${16^{**}}^{32}$	${18^{**}}^{30}$	${17^{*}}^{31}$	19^{29}	20^{28}	22^{26}	20^{28}	-	-	-	-	-
MxEU	${19^{*}}^{29}$	${21^{*}}^{27}$	20^{28}	19^{29}	20^{28}	21^{27}	26^{22}	27^{21}	-	-	-	-
αMM	${19^{*}}^{29}$	${20^{*}}^{28}$	20^{28}	20^{28}	22^{26}	23^{25}	22^{26}	23^{25}	24^{24}	-	-	-
CEV	${6^{***}}^{42}$	${7^{***}}^{41}$	${5^{***}}^{43}$	${9^{***}}^{39}$	${9^{***}}^{39}$	${10^{***}}^{38}$	${7^{***}}^{41}$	${7^{***}}^{41}$	${6^{***}}^{42}$	${6^{***}}^{42}$	-	-
SCEV	${1^{***}}^{47}$	${1^{***}}^{47}$	${1^{***}}^{47}$	${3^{***}}^{45}$	${3^{***}}^{45}$	${3^{***}}^{45}$	${2^{***}}^{46}$	${2^{***}}^{46}$	${2^{***}}^{46}$	${3^{***}}^{45}$	${4^{***}}^{44}$	-
EV	${2^{***}}^{46}$	${3^{***}}^{45}$	${2^{***}}^{46}$	${3^{***}}^{45}$	${3^{***}}^{45}$	${3^{***}}^{45}$	${1^{***}}^{47}$	${1^{***}}^{47}$	${2^{***}}^{46}$	${1^{***}}^{47}$	${2^{***}}^{46}$	${18^{*}}^{30}$

Table WA6: Winner counts with Wilkinson statistics (based on median predicted log-likelihood across 10 batches).

	SPT	SPT_u	SCEU	EU	MnEU	αMM	MxEU	CEU	GPT	GPT_u	CEV	EV
SPT_u	16^{32}_{*}	-	-	-	-	-	-	-	-	-	-	-
SCEU	18^{30}	23^{25}	-	-	-	-	-	-	-	-	-	-
EU	19^{29}_{*}	19^{29}	21^{27}	-	-	-	-	-	-	-	-	-
MnEU	18^{30}_{*}	19^{29}_{*}	20^{28}_{*}	21^{27}	-	-	-	-	-	-	-	-
αMM	18^{30}_{*}	17^{31}_{**}	19^{29}_{*}	18^{30}	22^{26}	-	-	-	-	-	-	-
MxEU	17^{31}_{*}	18^{30}_{*}	19^{29}	21^{27}_{*}	23^{25}	28^{20}	-	-	-	-	-	-
CEU	14^{34}_{***}	14^{34}_{***}	13^{35}_{***}	19^{29}_{*}	20^{28}_{*}	20^{28}	19^{29}	-	-	-	-	-
GPT	17^{31}_{***}	16^{32}_{***}	16^{32}_{***}	20^{28}_{*}	20^{28}_{*}	20^{28}_{*}	19^{29}_{*}	26^{22}	-	-	-	-
GPT_u	15^{33}_{***}	14^{34}_{***}	14^{34}_{***}	19^{29}_{**}	19^{29}_{**}	19^{29}_{**}	18^{30}_{**}	25^{23}	21^{27}	-	-	-
CEV	4^{44}_{***}	5^{43}_{***}	4^{44}_{***}	9^{39}_{***}	8^{40}_{***}	9^{39}_{***}	8^{40}_{***}	12^{36}_{***}	12^{36}_{***}	13^{35}_{**}	-	-
EV	4^{44}_{***}	3^{45}_{***}	3^{45}_{***}	2^{46}_{***}	3^{45}_{***}	3^{45}_{***}	3^{45}_{***}	5^{43}_{***}	6^{42}_{***}	6^{42}_{***}	6^{42}_{***}	-
SCEV	3^{45}_{***}	2^{46}_{***}	2^{46}_{***}	2^{46}_{***}	2^{46}_{***}	4^{44}_{***}	2^{46}_{***}	5^{43}_{***}	6^{42}_{***}	6^{42}_{***}	6^{42}_{***}	28^{20}

Table WA7: Winner counts with Wilkinson statistics (based on trimmed (0.01) mean of predicted log-likelihood across 10 batches).