Supplemental material

Controlling for presentation effects in choice

Yves Breitmoser
Bielefeld University

The supplementary material contains the robustness checks showing that the results obtain equally for alternative assumptions on subject heterogeneity and utility standardization (e.g. contextual utility, following Wilcox, 2011). The first sections review the econometric model and the experimental designs for ease of reference. The final section contains all parameter estimates including standard errors (for all models and all data sets).

Specifically, Section 1 provides all definitions required to replicate the econometric analysis. Section 2 provides additional information on the four analyzed data sets, including a full list of histograms. Section 3 analyzes subject heterogeneity and Section 4 analyzes utility standardization. Section 5 provides the robustness checks for the tests of in-sample accuracy and Section 6 provides respective goodness-of-fit plots and a first overview of parameter estimates. Section 7 provides robustness checks for counterfactual reliability, Section 8 provides those for estimate consistency. Section 9 provides the robustness checks for the analysis of the comparative relevance of controlling for subject heterogeneity and focality, and Section 10 provides those for capturing entropy. Section 11 finally provides the detailed list of parameter estimates.

Contents

1 Definition of the econometric models and measures

1.1 The models of stochastic choice considered

For convenience of the reader, this section reviews the models of stochastic choice considered in the analysis. For discussion of these models, let me refer to the paper.

Logit Given an option set X, a utility index $u : X \to \mathbb{R}$, and a degree of precision $\lambda \geq 0$, the logit choice probabilities $\Pr \in \Delta(X)$ are, for all $x \in X$,

$$\Pr(x) = \frac{\exp\{\lambda \cdot u(x)\}}{\sum_{x'} \exp\{\lambda \cdot u(x')\}}. \tag{1}$$

Focal Choice Adjusted Logit (FOCAL) FOCAL extends logit by allowing that choice probabilities also depend on the focality of the options under the assumption that focality is independent of utility. Given an option set X, a utility index $u : X \to \mathbb{R}$, a focality index $\phi : X \to \mathbb{R}$, a degree of precision $\lambda \geq 0$, and a degree of focality $\kappa \geq 0$, the FOCAL choice probabilities $\Pr \in \Delta(X)$ are, for all $x \in X$,

$$\Pr(x) = \frac{\exp\{\lambda\, u(x) + \kappa\, \phi(x)\}}{\sum_{x'} \exp\{\lambda \cdot u(x') + \kappa\, \phi(x')\}}. \tag{2}$$

Ordered GEV (OGEV, Small, 1987) Ordered GEV captures stochastic choice from ordered options, generalizing logit by allowing that the utility perturbations of proximate options (under the given ordering) are correlated. This correlation relaxes independence of irrelevant alternatives, is a special case of utility perturbations with generalized extreme value (GEV) distribution, and seems to capture choice patterns in dictator games more accurately than logit (Breitmoser, 2013). Given a utility function u, precision λ, degree of correlation κ, an integer bandwidth parameter $M < |X|$, and options represented by their integer ranks $s = 1, 2, \ldots$, the OGEV choice probabilities are

$$\Pr(s) = \sum_{r=s}^{s+M} \frac{w_{r-s} \exp\left\{\lambda u(s)/\kappa\right\}}{\exp\{I_r\}} \cdot \frac{\exp\{\kappa I_r\}}{\sum_{t=0}^{B+M} \exp\{\kappa I_t\}} \quad \text{with} \quad I_r = \ln \sum_{s' \in B_r} w_{r-s'} \exp\left\{\lambda u(s'|\alpha,\beta)/\kappa\right\}.$$

Perception Adjusted Luce Model (PALM) Echenique et al. (2014) introduce PALM to account for stochastic choice based on a weak perception ordering over options. Given a complete, transitive, and weak ordering $\succsim$ of options, the decision maker first considers the highest ranked options, potentially chooses one of them, otherwise process to the next layer of options, potentially chooses one of them, and so on. With positive probability, the decision maker does not consider all options in the process. PALM generalizes stochastic choice from consideration sets introduced by Manzini and Mariotti (2014) by allowing for a weak perception ordering. Manzini and Mariotti

(2014) in turn generalize Masatlioglu et al. (2012) by allowing for stochastic consideration sets. The weak perception ordering underlying PALM appears highly suitable to capture the focality effects observed in dictator choice. Echenique et al. (2014) define the Perception Adjusted Luce Model (PALM) without explicit restriction to logit choice, but facilitate comparisons with the other models considered here and, I focus on "perception adjustment" in relation to logit.

Given a utility function u, precision λ, and the perception ordering $\succsim$, the original PALM choice probabilities assume a specific degree of perception adjustment. I consider a generalized model that contains original PALM for $\kappa = 1$ as special case.

$$\tilde{\Pr}(x) = \mu(x,X) \cdot \prod_{X' \in X/\succsim : X' \succ x} \left(1 - \kappa \cdot \sum_{x' \in X : x' \in X'} \mu(x',X)\right) \qquad \text{with } \kappa \in [0,1] \tag{3}$$

with $\mu(x,X) = \text{Logit}(x) = \exp(\lambda\, u(x))/\sum_{x' \in X} \lambda\, u(x')$. Note that "$X/\succsim$" refers to the set of equivalence classes in which $\succsim$ partitions X, and that $X' \succ x$ holds if $x' \succ x$ for all $x' \in X'$.

The base probabilities $\tilde{\Pr}(x)$ add up to 1 only in the limiting logit model ($\kappa = 0$). Manzini and Mariotti (2014) and Echenique et al. (2014) allow for outside options to collect the remaining probability mass. There is no obvious outside option in dictator choice (for example, transferring "zero" is the primary option for many of the subjects). The natural assumption in this case (Manzini and Mariotti, 2014) appears to allow the decision maker to restart with the first level of options once the set of options is exhausted. This rescales the base probabilities such that they add up to 1, $\Pr(x) = \tilde{\Pr}(x)/\sum_{x'} \tilde{\Pr}(x')$.

1.2 Utility and utility standardization

Andreoni and Miller (2002), Fisman et al. (2007) and many subsequent analyses showed that utilities in standard Dictator games are in principle well-captured by utilities exhibiting constant elasticity of substitution (CES) over incomes. Using incomes (x_i, x_j), the utility of i is

$$u_i(x_i, x_j) = \big((1-\alpha)\cdot(1+x_i)^\beta + \alpha \cdot (1+x_j)^\beta\big)^{1/\beta}, \tag{4}$$

with α as degree of altruism and β as degree of efficiency concerns. Players focus on the lesser of the incomes and thus exhibit equity concerns if $\beta < 0$; in the extreme case $\beta \to -\infty$, preferences converge to Leontief. Players have efficiency concerns and focus on the sum of the incomes if $\beta \geq 1$; in the extreme case $\beta \to \infty$ they strictly maximize the larger of the incomes, which maximizes efficiency in the dictator games.

As is well known, if behavior exhibits random utility, choice probabilities are not invariant with respect to monotone transformations of the utility function. Specifically, it matters whether the incomes (x_i, x_j) in the CES utility function are the number of tokens of the players or the amount of money these tokens represent. I allow for either of these two models below, referring to them as **token-based** and **value-based** utilities, respectively. In addition, recent evidence suggests that subjects' utilities are normalized across conditions. For example, Wilcox (2011), Padoa-Schioppa (2009), and Padoa-Schioppa and Rustichini (2014); Rustichini and Padoa-Schioppa (2015) define the following two models to capture such standardization.

Contextual utility Wilcox (2011) defines contextual utility as the utility normalized to the range of utilities attainable in the current context. Given the utility function $u : X \to R$, the perceived utility is $\tilde{u}(x) = u(x) / \big(\max_{x'} u(x') - \min_{x'} u(x') \big)$. Note that, for motivation, Wilcox refers to "signal detection and stimulus discrimination experiments", similarly to Padoa-Schioppa and Rustichini (2014) and Rustichini and Padoa-Schioppa (2015).

Adaptive utility Padoa-Schioppa and Rustichini (2014) capture "adaptive coding" as a utility derived from the ranking of options. Given the utility function $u : X \to R$, the perceived utility is derived from the quantile function, $\tilde{u}(x) = \Pr\big(u(Y) \leq u(x)\big)$ where Y is uniform on $X \setminus \{x\}$.

In both models, utilities are normalized such that the difference between the maximal utility and the minimal utility in a given context is 1. The difference is whether the perceived utilities relate to the underlying absolute utilities or their ranking. Thus, both models appear to make very similar predictions, and for this reason, I will not explicitly consider both models. I focus on contextual utility, as this model is more closely related to the standard notions of token-based and value-based utilities.

1.3 Modeling heterogeneity

Each model considered here has a precision parameter λ, a degree of altruism α, and a degree of efficiency concerns β. All models but logit additionally have a "choice" parameter κ. In the analysis, I allow for heterogeneity, i.e. for these parameters to be distributed randomly across subjects. Using $\mathbf{p} = (\lambda, \kappa, \alpha, \beta)$ to describe the parameter profile of a given subject and $f(\cdot|\mathbf{d})$ to describe its joint density in the population given distribution parameters $\mathbf{d}$, the likelihood that the model with density $\mathbf{d}$ describes the choices o_s of subject $s \in S$ is

$$l(\mathbf{d}|o_s) = \int_{\mathbf{P}} f(\mathbf{p}|\mathbf{d}) \cdot \Pr(o_s|\mathbf{p})\, d\,\mathbf{p}, \tag{5}$$

with $\Pr(o_s|\mathbf{p})$ as the probability that o_s results under parameter profile $\mathbf{p}$ given the utility standardization and choice model in question. The integral is evaluated numerically using quasi-random numbers (described below). Aggregating over subjects, the log-likelihood of the model is

$$ll((\mathbf{d}|o) = \sum_{s \in S} \log l(\mathbf{d}|o_s) \tag{6}$$

with $o = \{o_s\}_{s \in S}$. A priori, experimental subjects are known to be heterogeneous in their preferences and in their precision when choosing optimally. I test the adequacy of this notion, and those of alternative models of heterogeneity, by investigating the following models.

Homogenous All subjects have the same parameter profile $(\lambda, \kappa, \alpha, \beta)$

Heterogenous Preferences The preference parameters (α, β) are heterogenous across subjects, the remaining parameters (precision and choice) are constant across subjects

Heterogenous Pref & Prec The preference parameters (α, β) and precision λ are heterogenous across subjects, the choice parameter is constant across subjects

Full heterogeneity All parameters $(\lambda, \kappa\alpha, \beta)$ are heterogenous across subjects

Heterogenous Pref & Choice The preference parameters (α, β) and choice κ are heterogenous across subjects, the precision parameter is constant across subjects

If a parameter is modeled to be heterogenous across subjects, then the parameter is assumed to be identically and independently distributed across subjects in the population. The respective distributions are standard:

Altruism α Normal with mean and variance as free parameters, truncated (not censored) to the interval $[-0.5, 0.5]$

Equity β Normal with mean and variance as free parameters

Precision λ Log-normal with mean and variance as free parameters

Choice κ Log-normal with mean and variance as free parameters, censored at upper bound 1 in case of OGEV and PALM

1.4 Computation and maximization of the log-likelihoods

The integral in Eq. (5) is evaluated by simulation, using quasi random numbers. This follows standard practice, see e.g. Train (2003). Parameters are estimated by maximizing the log-likelihood, sequentially applying two maximization algorithms. Initially I use the robust, gradient-free NEWUOA algorithm (Powell, 2006), and subsequently I verify convergence using a Newton-Raphson algorithm. The estimates are tested by extensive cross-analysis to ensure that global maxima are found. Specifically, for each model, I evaluate whether the estimates for the same model on other data sets fit better than the current estimates, whether the estimates on subsets of the same data set fit better than the current estimates, and whether the estimates obtained under other heterogeneity assumptions (suitably transformed) fit better than the current estimates. The estimation is considered "converged" once these cross-comparisons do not allow for further improvements of the log-likelihood. This generally took a fair number of "rounds" during which estimates and likelihoods kept improving. Thanks to the extent of these cross-comparisons, the robustness of the basic two-stage optimization procedure (Newuoa and Newton-Raphson), and manual inspection of the optimization paths, global maxima of the likelihood function seem to be obtained in all cases.

1.5 Evaluating model adequacy

Let $e \in E$ denote a given experiment. The data set associated with e is denoted D_e, the pooled data set is $D = \cup_e d_e$. The maximum likelihood estimate of the parameters given data set $D' \subset D$, i.e. the parameter vector defining ll as defined in Eq. (6), is denoted as $p(D')$.

Likelihood-based measures $LL(p|D')$ and $BIC(p|D')$ denote absolute log-likelihood and Bayes Information Criterion (Schwarz, 1978) associated with parameter vector p on data set D'. That is, given the log-likelihood ll as defined in Eq. (6),

$$LL(p|D') = \left|ll(p|D')\right| \quad BIC(p|D') = LL(p|D') + \log(\#\text{Observations}) \cdot \#\text{Parameters}/2, \tag{7}$$

using #Parameters as the number of free parameters fitted to the data set in question. By taking the absolute value of the log-likelihood, "less is better" for both of these likelihood-based measures. Note that considering the absolute value is without loss of information, as all choices are made from a discrete option sets and thus all basic log-likelihoods are negative.

Pseudo-R^2 The absolute difference between the log-likelihoods of two models are generally informative, but the actual scale of the log-likelihood is in itself uninformative. It is informative in relation to two benchmarks: the **minimal** log-likelihood $LL_{\min}$, which is the log-likelihood of the *naive model* predicting uniform randomization in all conditions, and the **maximal** log-likelihood $LL_{\max}$, which is the log-likelihood of the *clairvoyant model* predicting the observed relative frequencies in all conditions. The distance of a model's log-likelihood to $LL_{\min}$ is proportional to the variance explained by the model, the distance to $LL_{\max}$ is proportional to the variance left to be explained. In relation to both extremes, the Pseudo-R^2 (Nagelkerke, 1991)

$$\text{Pseudo-}R^2(p|D') = \frac{LL(p|D') - LL_{\min}(D')}{LL_{\max}(D') - LL_{\min}(D')} \tag{8}$$

measures the amount of variance explained by the model. In general, I will report both BIC and pseudo-R^2, to indicate both the absolute fit (which is the basis of model-discrimination) and the relative fit (which indicates where we stand in terms of model adequacy).

Counterfactual reliability The "counterfactual reliability" measures the accuracy of predictions using estimates from one experiment on data from another experiment. These other experiments represent "counterfactual scenarios" in this analysis. I analyze counterfactual reliability to three degrees: (i) reliability of preference estimates only, (ii) reliability of preference and precision estimates, and (iii) reliability of all estimates, i.e. preference, precision, and choice parameters. To facilitate formal definitions, I use the notation $\underset{P}{\wedge}$ and $\underset{PP}{\wedge}$. Namely, $p(D_{e'}) \underset{P}{\wedge} p(D_e)$ denotes the parameter vector that results if we use preference estimates from data set $D_{e'}$ and other parameter estimates from D_e, and $p(D_{e'}) \underset{PP}{\wedge} p(D_e)$ denotes the parameter vector that results if we use preference and precision estimates from $D_{e'}$ and other (if any) from D_e. Thus, the three degrees of counterfactual reliability can be defined as follows.

$$\text{Preferences:} \qquad CR_1(D_{e'}|D_e) = BIC\big(p(D_{e'}) \underset{P}{\wedge} p(D_e)|D_e\big) \tag{9}$$

$$\text{Pref \& Prec:} \qquad CR_2(D_{e'}|D_e) = BIC\big(p(D_{e'}) \underset{PP}{\wedge} p(D_e)|D_e\big) \tag{10}$$

$$\text{All estimates:} \qquad CR_3(D_{e'}|D_e) = LL\big(p(D_{e'})|D_e\big) \tag{11}$$

Note that the first two measures use parameters fitted to the data set in question (either precision and choice parameters or only choice parameters), i.e. the reference to BIC is not superfluous.

Estimate consistency The "estimate consistency" measures the difference between in-sample log-likelihood of a given model and the out-of-sample log-likelihood of the same model using estimates from a different experiments. These differences are the basis of tests whether parameter estimates are transferable across experiments, and presentation effects thus factored out, or significantly different between experiments and thus appear context dependent. Mirroring the analysis of counterfactual reliability, I focus on consistency of either preference estimates on their own, preference and precision estimates, or all estimates.

$$\text{Preferences:} \qquad EC_1(D_{e'}|D_e) = BIC\big(p(D_{e'}) \underset{P}{\wedge} p(D_e)|D_e\big) - BIC\big(p(D_e)|D_e\big) \tag{12}$$

$$\text{Pref \& Prec:} \qquad EC_2(D_{e'}|D_e) = BIC\big(p(D_{e'}) \underset{PP}{\wedge} p(D_e)|D_e\big) - BIC\big(p(D_e)|D_e\big) \tag{13}$$

$$\text{All estimates:} \qquad EC_3(D_{e'}|D_e) = LL\big(p(D_{e'})|D_e\big) - BIC\big(p(D_e)|D_e\big) \tag{14}$$

1.6 Standard errors and model selection

When estimating standard errors and testing likelihood ratios, I account for the possibility that models may be misspecified (which is the implicit hypothesis of the analysis for all models but FOCAL). Specifically, I use Huber Sandwich Estimators of standard errors and likelihood-ratio tests following Schennach and Wilhelm (2016), see below.

Throughout the paper and the supplementary material, I indicate significance (p-values) at two levels, "weak significance" for $p = 0.05$ and "high significance" for $p = 0.005$. The former standard level has limited relevance in the present analysis, as I generally conduct multiple tests simultaneously—across models and measures of goodness of fit—to avoid the reliance on a single p-value and document the "big picture" and robustness. There tend to be around 6-16 tests run simultaneously depending on the question to be analyzed, and the respective Bonferroni correction is implemented approximately at $p = 0.005$. Addressing the multiple testing problem, I consider $p = 0.005$ the basic level required for significance and report "weak significance" at the 0.05 level for transparency.

Standard errors I use the Huber Sandwich Estimator of standard errors, which is robust to model misspecification and allows to capture the panel structure of the data (for discussion, see Freedman, 2012). As before, subjects are indexed by $s = 1, \dots, n$, the observations regarding subject s are denoted by o_s, and the vector of parameters is denoted as $\mathbf{d}$. The individual likelihood $l_s(\mathbf{d})$ is the probability that model $\mathbf{d}$ generates observations o_s. Using this notation, the aggregate log-likelihood is

$$ll(\mathbf{d}) = \sum_s ll_s(\mathbf{d}) = \sum_s \log l_s(\mathbf{d}). \tag{15}$$

The derivatives for ll_s for $i = 1, \dots, n$ are

$$g_s(\mathbf{d}) = \frac{\partial ll_s(\mathbf{d})}{\partial \mathbf{d}} = \nabla_{\mathbf{d}}\, ll_s(\mathbf{d}). \tag{16}$$

The variance-covariance matrix of ll at the MLE d is estimated as

$$\hat{E} = (-H)^{-1}\, V\, (-H)^{-1} \qquad \text{with} \quad H = \nabla^2_{\mathbf{d}}(\mathbf{d}) \quad \text{and} \quad V = \sum_s g_s(\mathbf{d})^T g_s(\mathbf{d}) \tag{17}$$

The square roots of the diagonal elements are the standard errors.

Likelihood ratio tests The likelihood ratio test follows Schennach and Wilhelm (2016), as this test tractably provides robustness to misspecification and a uniform procedure for nested, non-nested, and overlapping model structures. The models are denoted A and B, and the likelihood-ratio of interest is the difference between the log-likelihoods of models A and B. First, the likelihood ratio is weighted,

$$lr = \frac{1}{n}\sum_s \big(\omega_s(\varepsilon) ll_{A,i}(\mathbf{d}_A) - \omega_{i+1}(\varepsilon) ll_{B,i}(\mathbf{d}_B)\big), \tag{18}$$

where ll_A and ll_B are the models' log-likelihoods, with weights (for $j = 1, \dots, n+1$)

$$\omega_j(\varepsilon) = \begin{cases} 1, & k \text{ odd}, \\ 1+\varepsilon, & k \text{ even}, \end{cases} \tag{19}$$

Here, ε is called regularization parameter, and $\varepsilon = 0$ yields the standard non-nested Vuong test. Schennach and Wilhelm (2016) generally recommend ε to be

$$\varepsilon = c_\alpha n^{-1/4}\sqrt{\ln\ln n} \qquad \text{with} \quad c_\alpha = \sqrt{\frac{4\phi(z_{1-\alpha})\Lambda}{\phi(z_b)(z_b+z_{1-\alpha})(\sigma_A^2+\sigma_B^2)}}$$
$$\Lambda = \max\left\{\left|tr(H_A^{-1}V_A)\right|, \left|tr(H_B^{-1}V_B)\right|\right.$$
$$z_b = \frac{1}{2}\left(-z_{1-\alpha} + (4+z_{1-\alpha}^2)^{1/2}\right)$$

with $z_{1-\alpha}$ as $(1-\alpha)$-quantile of the standard normal distribution and ϕ as its density, H and V as the matrices defined above, and σ_A^2 and σ_B^2 as the sample variances (of the log-likelihoods across individuals in the two samples). Next, define σ_{AB} as the covariance of the sample likelihoods. The variance of the likelihood ratio is $\sigma^2 = \sigma_A^2 - \sigma_{AB} + \sigma_B^2$ in the uniformly weighted case ($\varepsilon = 0$). In the reweighted case ($\varepsilon > 0$), the test statistic

$$t = \frac{\sqrt{n}\cdot lr}{\tilde\sigma} \qquad \text{with} \quad \tilde\sigma^2 = (1+\varepsilon)\sigma^2 + \frac{\varepsilon^2}{2}(\sigma_A^2+\sigma_B^2), \tag{20}$$

where the $\tilde\sigma^2$ is the sample variance of the weighted likelihood ratio, is approximately normal under H_0. The authors indicate that $\varepsilon = 0.5$ and $\varepsilon = 1$ are generally reasonable substitutes for the data-set specific recommendation of ε. More generally, this implies that the above definition of ε using $c_\alpha = 1$ is a good approximation for the optimal value of ε: it responds to the size of the data set in the "optimal way" and tends to be in the interval indicated by Schennach and Wilhelm. For these reasons, I use Schennach-Wilhelm tests with $c = 1$.

References

Andreoni, J. and Miller, J. (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753.

Breitmoser, Y. (2013). Estimation of social preferences in generalized dictator games. *Economics Letters*, 121(2):192–197.

Cappelen, A., Hole, A., Sørensen, E., and Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review*, 97(3):818–827.

Echenique, F., Saito, K., and Tserenjigmid, G. (2014). The perception-adjusted luce model. *Working paper*.

Fisman, R., Kariv, S., and Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, 97(5):1858–1876.

Freedman, D. A. (2012). On the so-called "huber sandwich estimator" and "robust standard errors". *The American Statistician*, pages 299–302.

Harrison, G. W. and Johnson, L. T. (2006). Identifying altruism in the laboratory. In Isaac, R. M. and Davis, D. D., editors, *Experiments Investigating Fundraising and Charitable Contributors*, volume 11 of *Research in experimental economics*, pages 177–223. Emerald Group Publishing Limited.

Manzini, P. and Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, 82(3):1153–1176.

Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2012). Revealed attention. *American Economic Review*, 102(5):2183–2205.

Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692.

Padoa-Schioppa, C. (2009). Range-adapting representation of economic value in the orbitofrontal cortex. *Journal of Neuroscience*, 29(44):14004–14014.

Padoa-Schioppa, C. and Rustichini, A. (2014). Rational attention and adaptive coding: a puzzle and a solution. *American Economic Review*, 104(5):507.

Powell, M. (2006). The newuoa software for unconstrained optimization without derivatives. *Large-Scale Nonlinear Optimization*, pages 255–297.

Rustichini, A. and Padoa-Schioppa, C. (2015). A neuro-computational model of economic decisions. *Journal of Neurophysiology*, 114(3):1382–1398.

Schennach, S. and Wilhelm, D. (2016). A simple parametric model selection test. *Journal of the American Statistical Association*.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464.

Small, K. (1987). A discrete choice model for ordered alternatives. *Econometrica*, 55(2):409–424.

Train, K. (2003). *Discrete choice methods with simulation*. Cambridge Univ Pr.

Wilcox, N. (2011). Stochastically more risk averse: A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics*, 162(1):89–104.

2 Overview of experimental designs and choice patterns

Tables 3 and 5 provide further details on tables contained in the paper, by providing the information separately for each treatment (in each experiment). Table 3 provides detailed information on the designs of all treatments, Table 5 describes the distribution of choices across "round" numbers in each treatment. Tables 1 and 6 provide summary statistics about the distribution of options chosen, such as entropy and related measures, relative frequency of the equal split, and relative frequency of top-level options, once aggregated across treatments (per experiment) and once per treatment and experiment. Table 2 reviews the budgets and their monetary values across experiments, and Table 7 breaks it down to the treatment level. Figures 1–2 provides histograms of the choices for each treatment in each experiment (aggregating appropriately in FKM07).

Table 1: Characterizing distribution and clustering of the choices (averages across treatments)

	Equal split			Uniformity				Top Level of rounding		
Treat	Label	Rel Lev	Rel Freq	Q(100)	Q(90)	Entropy	exp(Entr)	#Options	Rel Freq	Other
AM02	10–50	-1,0	0.16–0.36	15.62	6.38	1.78	6.08	5–11	0.82	0.18
HJ06	8–37	-1,-2,0	0.02–0.39	15.3	10.3	2.37	11.08	5–11	0.67	0.33
FKM07	NA–NA	-1	0.01–0.03	54.62	37.5	3.41	30.58	16.7–66.3	0.24	0.76
CHST07	200–800	0	0.16–0.67	6.5	4.67	1.64	5.49	5–17	0.98	0.02

Note: The tables provides further information on the experimental designs and dictator behavior.
Columns on "Equal Split" The "label" gives the number to be entered to induce the equal split (the range across treatments). "RelLev" gives the relative level of rounding, i.e. the number of levels of roundness that the equal split is below the "Top level" defined below. "RelFreq" lists the relative frequency of the equal split.
Columns on "Uniformity" $Q(100)$ is the number of different options chosen, $Q(90)$ is the (minimal) number of options accounting for 90% of the choices. "Entropy" gives the Shannon entropy and "exp(Entropy)" is its reversion back to the "number of options considered".
Columns on "Top level of rounding" The top level of rounding is, if all options are ordered by their level of rounding, given by the third-highest option in this ranking. The option to transfer "zero" also is considered to be at the top level of rounding. In FKM07, choosing a number at the top level is equated with choosing an integer, which virtually always means choosing the minimal option. "RelFreq" lists the relative frequency of choosing options at the top level of rounding, "Other" lists the relative frequency of choosing other options.

Table 2: Budgets and their monetary values in the analyzed experiments

	#Tokens	Value for D	Value for R	Prob(Relevant)
AM02	40–100	4–15	4–15	1/8
HJ06	40-100	8–20	15–32	1/10
CHST07	400–1600	67–267	67–267	1/2

Note: "#Tokens" provides the range of the number of tokens to be redistributed by the Dictator (across treatments in each experiment), "Value for D" provides the respective range of values of the Dictator in US Dollar, "Value for R" provides the respective range of values for the Recipient, and "Prob(Relevant)" provides the probability that a given decision of the Dictator is payoff relevant. FKM07 is skipped, as treatments as such are not defined.

Table 3: Information on the various experiments and treatments

(a) Dictator games I (AM02)

Treat	#Options	#Obs	Ratio	Mean
1	41	176	3:1	8.02
2	41	176	1:3	12.81
3	61	176	2:1	12.68
4	61	176	1:2	19.4
5	76	176	2:1	15.51
6	76	176	1:2	22.68
7	61	176	1:1	14.55
8	101	176	1:1	23.03

(b) Dictator games II (HJ06)

Treat	#Options	#Obs	Ratio	Mean
1	76	57	2:4	25.12
2	41	57	2:6	12.54
3	76	57	2:2	20.11
4	61	57	2:4	20.56
5	41	57	5:5	11
6	101	57	1:2	32.86
7	61	57	2:3	19.93
8	81	57	2:4	25.88
9	41	57	2:5	12.3
10	41	57	2:8	12.25

(c) Dictator games with GUI (FKM07)

Treat	#Options	#Obs	Ratio	Mean
1	NA	96	$\approx$ 4:1	20.58
2	NA	160	$\approx$ 3:1	22.72
3	NA	168	$\approx$ 5:2	20.71
4	NA	199	$\approx$ 2:1	21.73
5	NA	261	$\approx$ 3:2	19.74
6	NA	393	$\approx$ 1:1	18.68
7	NA	176	$\approx$ 1:2	18.81
8	NA	159	$\approx$ 1:3	15.79

(d) Dictator games III (CHST07)

Treat	#Options	#Obs	Ratio	Mean
1	401	18	1:1	155.56
2	601	64	1:1	147.66
3	801	38	1:1	196.05
4	1001	12	1:1	325
5	1201	38	1:1	389.47
6	1601	20	1:1	260

Note: These tables list detailed information on all treatments in all experiments: the number of options, the number of observations, the transfer ratio, and the mean transfer. Regarding FKM07, as the transfer ratios and budget sizes (in tokens) are randomized across the 50×76 decisions, well-defined treatments applying to all subjects do not exist. In this table, I focus on the subsets of the data that are "close to" to conventional transfer ratios: A given transfer ratio is consider "close to" any of the conventional ones if it does not differ by more than 10 percent from it. in FKM07, the mean transfer is the mean share of the dictator budget transferred.

Table 4: Determinants of the "number of options considered"

Number of options considered against	All	AM02	HJ06	CHST07
Number of options available	-0.16	0.32	0.76^{-}	0.66
Number of options at top level	0.16	0.32	0.76^{-}	0.66
Relative frequency of top level	-0.73^{--}	-0.64	-0.42	-0.28
Focality level of Equal Split	-0.78^{--}	-0.41	-0.73^{-}	*NA*
Relative frequency of Equal Split	-0.61^{--}	0.21	-0.67^{-}	-0.94^{-}
Number of options comprising 90%	0.95^{--}	0.88^{--}	0.93^{--}	0.97^{--}

Note: This table provides Spearman coefficients of rank correlation of "exp(Entr)" (the exponent of the entropy) in a given treatment with various statistics in the respective treatment. Significance at .005 is indicated by $^{--}$, and significance at .05 is indicated by $^{-}$.

Table 5: Distribution of choices across "round" numbers by treatment

	Strategies		Percentage of choices with greatest factor …												
Treat	Min	Max	0	1000	500	250	100	50	25	10	5	2.5	1	0.5	other
Dictator games I (Andreoni and Miller, 2002): AM02															
1	0	40	**56**						1	**31**	7	0	6		
2	0	40	**30**						2	**61**	5	0	3		
3	0	60	**49**					1	1	**39**	6	0	4		
4	0	60	**25**					2	1	**59**	9	0	6		
5	0	75	**52**					16	11	6	11	0	3		
6	0	75	**27**					7	**41**	11	10	0	4		
7	0	60	**41**					0	0	**51**	5	0	4		
8	0	100	**45**				1	**31**	5	14	2	0	2		
All			**39**				1	9	7	**33**	6	0	4		
Dictator games II (Harrison and Johnson, 2006): HJ06															
1	0	75	**23**					2	21	28	21	0	5		
2	0	40	**19**						0	**53**	18	0	11		
3	0	75	**30**					2	7	14	**32**	0	16		
4	0	60	**21**					2	2	**60**	12	0	4		
5	0	40	**39**						4	**44**	7	0	7		
6	0	100	**21**				4	12	9	**37**	4	0	14		
7	0	60	**18**					2	12	**47**	11	0	11		
8	0	80	**19**					5	16	**44**	4	0	12		
9	0	40	**23**						2	**49**	12	0	14		
10	0	40	**21**						2	**35**	26	0	16		
All			**22**				3	4	7	**39**	14	0	10		
Dictator games with GUI (Fisman et al., 2007): FKM07															
1	0	100	**25**				0	0	0	2	3	2	2	5	**60**
2	0	100	**26**				0	0	0	1	2	2	2	6	**62**
3	0	100	**24**				0	0	0	0	1	1	5	5	**64**
4	0	100	**21**				0	0	0	0	2	1	4	6	**67**
5	0	100	**28**				0	0	0	1	1	0	4	5	**61**
6	0	100	**30**				0	0	0	1	2	1	5	4	**58**
7	0	100	**19**				0	0	0	0	3	1	3	4	**70**
8	0	100	**25**				0	0	0	0	1	1	4	7	**64**
All			**25**				0	0	0	0	2	1	4	5	**63**
Dictator games III (Cappelen et al., 2007): CHST07															
1	0	400	**22**			0	**78**								
2	0	600	**31**		0	2	**64**	3							
3	0	800	**32**		3	0	**63**	3							
4	0	1000	**25**	0	17	0	**58**								
5	0	1200	**24**	0	8	3	**63**	3							
6	0	1600	**50**	0	0	0	**50**								
All			**30**	0	5	1	**62**	1	0	0	0	0	0		

Table 6: Characterizing distribution and clustering of the choices by treatment

Treat	Equal split			Uniformity				Top Level of rounding		
	Label	Rel Lev	Rel Freq	Q(100)	Q(90)	Entropy	exp(Entr)	#Options	Rel Freq	Other
Dictator games I (Andreoni and Miller, 2002): AM02										
1	30	0	0.16	12	5	1.51	4.54	5	0.87	0.13
2	10	0	0.36	11	5	1.74	5.67	5	0.91	0.09
3	40	0	0.17	16	6	1.74	5.67	7	0.89	0.11
4	20	0	0.31	19	9	2.15	8.62	7	0.85	0.15
5	50	0	0.16	17	7	1.75	5.75	8	0.74	0.26
6	25	-1	0.35	21	8	2.09	8.11	8	0.45	0.55
7	30	0	0.34	14	5	1.61	5.01	7	0.91	0.09
8	50	0	0.31	15	6	1.66	5.24	11	0.91	0.09
Dictator games II (Harrison and Johnson, 2006): HJ06										
1	25	-1	0.19	16	11	2.52	12.45	8	0.53	0.47
2	10	0	0.35	13	8	2.15	8.59	5	0.72	0.28
3	37	-2	0.07	15	10	2.39	10.88	8	0.46	0.54
4	20	0	0.28	13	8	2.21	9.14	7	0.82	0.18
5	20	0	0.39	11	6	1.65	5.19	5	0.82	0.18
6	33	-2	0.09	17	12	2.65	14.21	11	0.74	0.26
7	24	-2	0.02	18	13	2.57	13.12	7	0.67	0.33
8	27	-2	0.02	19	14	2.73	15.36	9	0.68	0.32
9	11	-2	0.02	16	11	2.36	10.56	5	0.72	0.28
10	8	-2	0.05	15	10	2.43	11.33	5	0.56	0.44
Dictator games with GUI (Fisman et al., 2007): FKM07										
1	NA	-1	0.03	43	34	3.28	26.67	16.7	0.3	0.7
2	NA	-1	0.02	58	42	3.54	34.61	22	0.2	0.8
3	NA	-1	0.01	56	40	3.53	34.05	26.1	0.22	0.78
4	NA	-1	0.02	55	37	3.55	34.93	33.5	0.19	0.81
5	NA	-1	0.01	61	41	3.44	31.26	44.2	0.26	0.74
6	NA	-1	0.02	63	37	3.33	27.84	64.1	0.24	0.76
7	NA	-1	0.03	53	36	3.4	29.97	66.3	0.25	0.75
8	NA	-1	0.01	48	33	3.23	25.31	64.6	0.3	0.7
Dictator games III (Cappelen et al., 2007): CHST07										
1	200	0	0.67	4	3	1.01	2.74	5	1	0
2	300	0	0.28	6	4	1.55	4.71	7	0.95	0.05
3	400	0	0.26	7	5	1.71	5.55	9	0.97	0.03
4	500	0	0.17	6	5	1.91	6.77	11	1	0
5	600	0	0.16	10	7	2.14	8.46	13	0.95	0.05
6	800	0	0.2	6	4	1.55	4.73	17	1	0

Note: See Table 1.

Table 7: Budgets and their monetary values by treatment

(a) Dictator games I (AM02)

Treatment	#Tokens	Value for D	Value for R	Prob(Relevant)
1	40	12	4	$1/8$
2	40	4	12	$1/8$
3	60	12	6	$1/8$
4	60	6	12	$1/8$
5	75	15	7.50	$1/8$
6	75	7.50	15	$1/8$
7	60	6	6	$1/8$
8	100	10	10	$1/8$

(b) Dictator games II (HJ06)

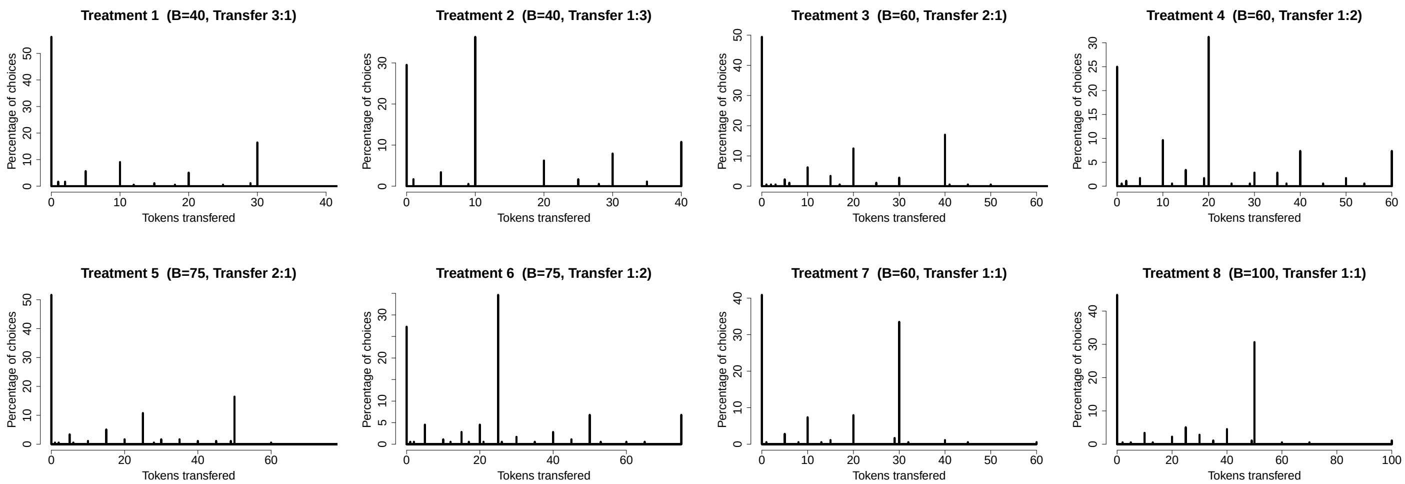
Treatment	#Tokens	Value for D	Value for R	Prob(Relevant)
1	75	15	30	$1/10$
2	40	8	24	$1/10$
3	75	15	15	$1/10$
4	60	12	24	$1/10$
5	40	20	20	$1/10$
6	100	10	20	$1/10$
7	60	12	18	$1/10$
8	80	16	32	$1/10$
9	40	8	20	$1/10$
10	40	8	32	$1/10$

(c) Dictator games III (CHST07)

Treatment	#Tokens	Value for D	Value for R	Prob(Relevant)
1	400	67	67	$1/2$
2	600	100	100	$1/2$
3	800	133	133	$1/2$
4	1000	167	167	$1/2$
5	1200	200	200	$1/2$
6	1600	267	267	$1/2$

Figure 1: Experimental observations: Standard Dictator games (AM02 and HJ06)

(a) Dictator games I (Andreoni and Miller, 2002): AM02



(b) Dictator games II (Harrison and Johnson, 2006): HJ06

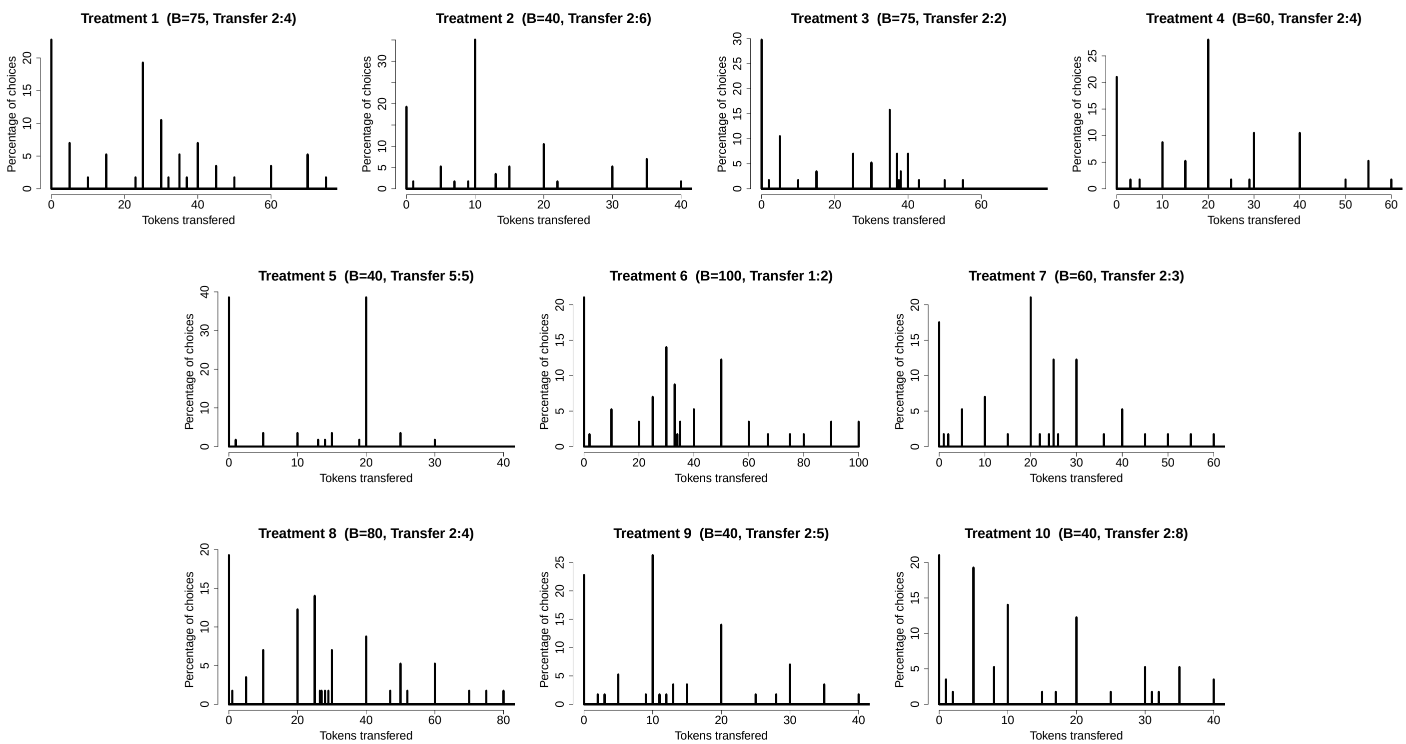
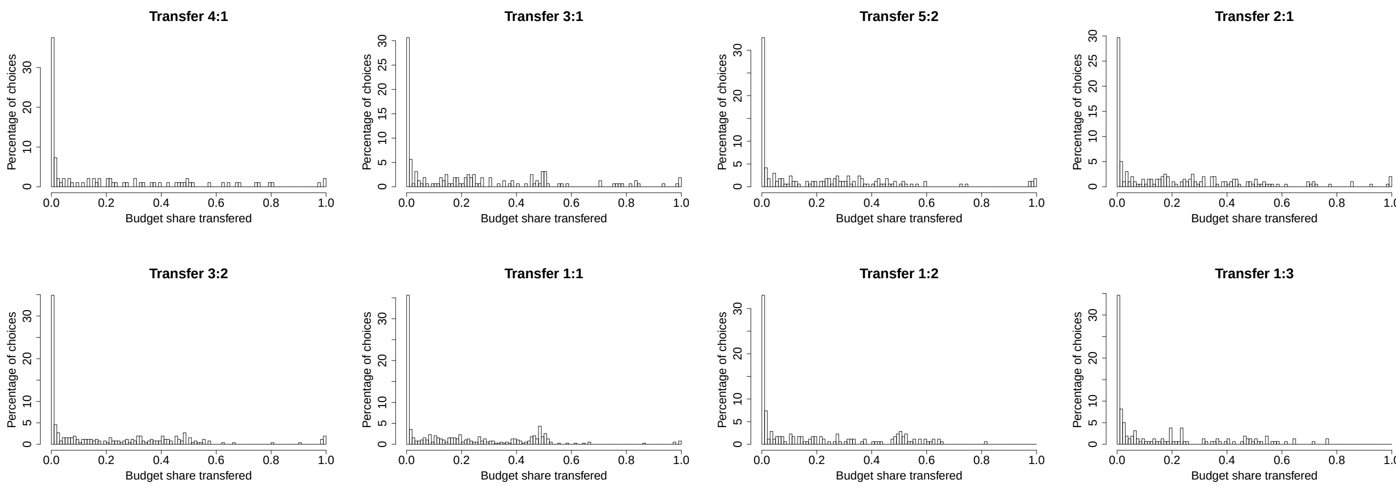
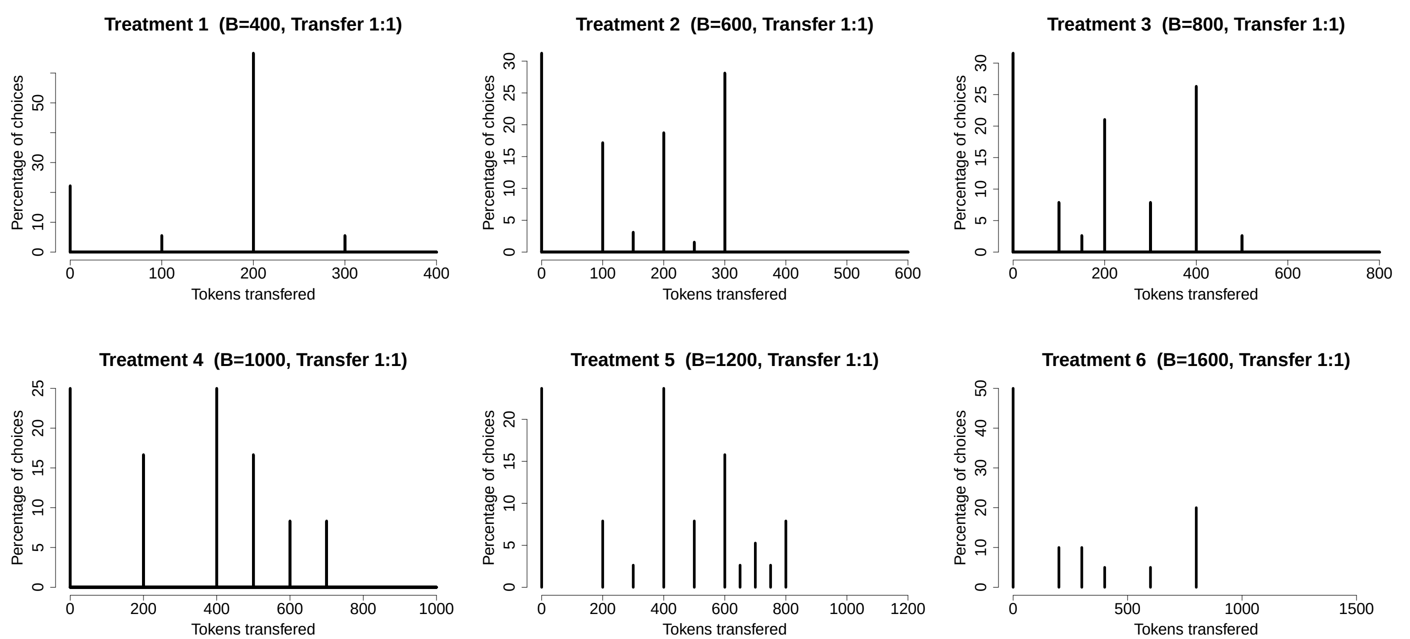


Figure 2: Experimental observations: Non-standard dictator games (FKM07 and CHST07)

(a) Dictator games with GUI (Fisman et al., 2007): FKM07



(b) Dictator games III (Cappelen et al., 2007): CHST07



3 How do subjects differ?

This section analyzes the extent of subject heterogeneity across data sets, to identify the most adequate econometric model. The main result is that, taking heterogeneity of preferences as given (based on the existing literature), the precision parameter is significantly heterogeneous, and given this, heterogeneity of the choice bias κ is not significant but tends to improve predictive adequacy across data sets and models (see also the next section).

Specifically, for all choice models (Logit, OGEV, PALM, FOCAL), all forms of utility standardization, and all data sets, the following can be observed: (i) heterogeneity of both preference and precision generally improves upon heterogeneity solely of preferences, mostly highly significantly; (ii) heterogeneity of preferences, precision and choice generally improves upon heterogeneity of preferences and choice, mostly highly significantly; and (iii) heterogeneity of preferences, precision and choice improves on heterogeneity of preferences and precision, both in-sample and out-of-sample, but rarely significantly so (see also next section). Heterogeneity of precision follows from (i) and (ii), and given this heterogeneity of choice is insignificant but in its tendency robust.

Table 8 provides an overview of these relations in-sample, pooling all data, and Tables 9–12 provide the results for each data set individually.

Table 8: How do subjects differ? All experiments pooled

(a) Value-based utility

Model	Heterogeneity of ... Preferences		Pref & Prec		Full		Pref & Choice
Logit	8889.1	$\gg$	7735	$\approx$	7730.3	$\ll$	8893.8
FOCAL	6453	$>$	5839.2	$\approx$	5813.4	$<$	6351.7
PALM	8592.7	$\gg$	7516.1	$\approx$	7474.2	$\ll$	8593.1
OGEV	8136.1	$>$	7388.8	$\approx$	7334.6	$\ll$	8229.3

(b) Token-based utility

Model	Heterogeneity of ... Preferences		Pref & Prec		Full		Pref & Choice
Logit	8981.4	$\gg$	7651.6	$\approx$	7650.5	$\ll$	8987
FOCAL	6543.7	$>$	5830.6	$\approx$	5811.4	$\ll$	6467
PALM	8742.7	$\gg$	7445.7	$\approx$	7427.8	$\ll$	8686.8
OGEV	8410.7	$\gg$	7363.6	$\approx$	7333.8	$\ll$	8299.7

(c) Contextual utility

Model	Heterogeneity of ... Preferences		Pref & Prec		Full		Pref & Choice
Logit	8973.1	$\gg$	7728	$\approx$	7710	$\ll$	9003.7
FOCAL	6528.7	$>$	5901.2	$\approx$	5849.6	$<$	6453.3
PALM	8660.4	$\gg$	7526.2	$\approx$	7466.3	$\ll$	8640.7
OGEV	8250.2	$\gg$	7432.9	$\approx$	7332.9	$\ll$	8068

Note: This table comprises three panels, covering the different standardizations of tokens, and each panel provides measures of the descriptive adequacy (in-sample BIC) for 4×4 models: the four models of heterogeneity and the four models of choice. Throughout, BIC is measured as distance to the log-likelihood of clairvoyance (less is better), $\approx$ indicates p-values above 0.05, $>, <$ indicate p-values between 0.005 and 0.05, and $\gg, \ll$ indicate p-values below 0.005. In all cases, the likelihood-ratio tests follow Schennach and Wilhelm (2016).

Table 9: How do subjects differ? Analysis of AM02

(a) Value-based utility

Model	Heterogeneity of … Preferences		Pref & Prec		Full		Pref & Choice
Logit	1710.2	$\gg$	1241.7	$\approx$	1241.2	$\ll$	1709
FOCAL	394.1	$>$	339.6	$\approx$	318.5	$\approx$	377.3
PALM	1513	$\gg$	1107.3	$\approx$	1082.8	$\ll$	1520.3
OGEV	1128.5	$\gg$	964	$\approx$	958.2	$\ll$	1296.3

(b) Token-based utility

Model	Heterogeneity of … Preferences		Pref & Prec		Full		Pref & Choice
Logit	1754.6	$\gg$	1201.3	$\approx$	1200.4	$\ll$	1750
FOCAL	402.6	$>$	321	$\approx$	317.8	$\approx$	393.3
PALM	1603.3	$\gg$	1066.9	$\approx$	1065.1	$\ll$	1549
OGEV	1309.9	$\gg$	966.6	$\approx$	947.6	$\ll$	1292

(c) Contextual utility

Model	Heterogeneity of … Preferences		Pref & Prec		Full		Pref & Choice
Logit	1753.2	$\gg$	1264.5	$\approx$	1262.6	$\ll$	1777.4
FOCAL	445.7	$>$	348.5	$\approx$	348	$\approx$	419.8
PALM	1591.4	$\gg$	1113.3	$\approx$	1095.9	$\ll$	1578.7
OGEV	1161.9	$\gg$	984.1	$\approx$	954.4	$\ll$	1273

Note: The layout is equivalent to Table 8, but the focus is on the data of AM02.

Table 10: How do subjects differ? Analysis of HJ06

(a) Value-based utility

	Heterogeneity of ...						
Model	Preferences		Pref & Prec		Full		Pref & Choice
Logit	804	$\approx$	703.1	$\approx$	703.1	$\approx$	804.1
FOCAL	177.8	$\approx$	168.5	$\approx$	168.4	$\approx$	180.8
PALM	779.4	$\approx$	681.2	$\approx$	681.2	$\approx$	775
OGEV	747.2	$\approx$	671.7	$\approx$	692.4	$\approx$	804.1

(b) Token-based utility

	Heterogeneity of ...						
Model	Preferences		Pref & Prec		Full		Pref & Choice
Logit	792.1	$>$	670.3	$\approx$	670.3	$<$	793.8
FOCAL	179.6	$\approx$	169.7	$\approx$	169.6	$\approx$	182.3
PALM	768.5	$>$	649.7	$\approx$	649.7	$<$	765.6
OGEV	733.2	$>$	641.4	$\approx$	659.1	$<$	784.8

(c) Contextual utility

	Heterogeneity of ...						
Model	Preferences		Pref & Prec		Full		Pref & Choice
Logit	794.2	$>$	669.7	$\approx$	669.6	$<$	791.9
FOCAL	165.5	$\approx$	166.7	$\approx$	166.7	$\approx$	168.2
PALM	763.5	$>$	649.1	$\approx$	649.1	$\approx$	754
OGEV	748.3	$>$	636.9	$\approx$	658.3	$<$	791.9

Note: The layout is equivalent to Table 8, but the focus is on the data of HJ06.

Table 11: How do subjects differ? Analysis of FKM07

(a) Value-based utility

	Heterogeneity of ...						
Model	Preferences		Pref & Prec		Full		Pref & Choice
Logit	5632.7	≈	5111.1	≈	5107	≈	5636.6
FOCAL	5599.7	≈	5089.6	≈	5089.1	≈	5518
PALM	5620.9	≈	5099	≈	5093.5	≈	5617.1
OGEV	5551.2	≈	5084.3	≈	5025.8	≈	5441.9

(b) Token-based utility

	Heterogeneity of ...						
Model	Preferences		Pref & Prec		Full		Pref & Choice
Logit	5689.1	≈	5102.2	≈	5102.1	<	5695.8
FOCAL	5675.8	≈	5093.7	≈	5082	<	5611.9
PALM	5687.6	≈	5098.3	≈	5097.6	<	5687.5
OGEV	5651.4	≈	5080.5	≈	5065.9	≈	5535

(c) Contextual utility

	Heterogeneity of ...						
Model	Preferences		Pref & Prec		Full		Pref & Choice
Logit	5689.3	≈	5118.2	≈	5102.1	<	5695.8
FOCAL	5630.1	≈	5143.1	≈	5099.9	≈	5584.2
PALM	5630.1	≈	5144.8	≈	5107.1	<	5631.8
OGEV	5629.6	≈	5138.9	≈	5059.1	≈	5314.9

Note: The layout is equivalent to Table 8, but the focus is on the data of FKM07.

Table 12: How do subjects differ? Analysis of CHST07

(a) Value-based utility

Model	Heterogeneity of …						
	Preferences		Pref & Prec		Full		Pref & Choice
Logit	804	≈	703.1	≈	703.1	≈	804.1
FOCAL	177.8	≈	168.5	≈	168.4	≈	180.8
PALM	779.4	≈	681.2	≈	681.2	≈	775
OGEV	747.2	≈	671.7	≈	692.4	≈	804.1

(b) Token-based utility

Model	Heterogeneity of …						
	Preferences		Pref & Prec		Full		Pref & Choice
Logit	792.1	>	670.3	≈	670.3	<	793.8
FOCAL	179.6	≈	169.7	≈	169.6	≈	182.3
PALM	768.5	>	649.7	≈	649.7	<	765.6
OGEV	733.2	>	641.4	≈	659.1	<	784.8

(c) Contextual utility

Model	Heterogeneity of …						
	Preferences		Pref & Prec		Full		Pref & Choice
Logit	794.2	>	669.7	≈	669.6	<	791.9
FOCAL	165.5	≈	166.7	≈	166.7	≈	168.2
PALM	763.5	>	649.1	≈	649.1	≈	754
OGEV	748.3	>	636.9	≈	658.3	<	791.9

Note: The layout is equivalent to Table 8, but the focus is on the data of CHST07.

4 Are utility scales adaptive?

This section analyzes "utility standardization" as described in Section 1.2. For convenience, let me remind that token-based utility assumes the "incomes" in the CES utility function are the tokens received by the players, value-based utility assumes the "incomes" are the respective monetary terms in US Dollar, and contextual utility normalizes the raw utility such that it is bounded by 0 and 1 in each choice problem.

Starting point of the analysis is the observation from the previous section that subjects are heterogeneous in their precision (besides preferences), and tend to be heterogeneous also in their choice bias κ. This gives two potential models of heterogeneity, in relation to which utility standardization is evaluated.

The main result is that there are no differences in-sample that are significant in pairwise tests (Table 13) and few out-of-sample that are significant in pairwise tests (Table 14). Tables 15–18 provide the respective analyses for each data set. This shows that the choice between token-based, contextual, and value-based models will not be consequential with respect to the main results. To clarify this, however, all subsequent results are reported for all approaches to utility standardization as robustness checks. Further, the analysis confirms that the models with full heterogeneity tend to be more adequate than the models assuming heterogeneity only of preferences and precision (i.e. not of choice bias κ). Given this, in relation to Logit and OGEV, token-based utility tend to be most adequate, while value-based utility and contextual utility tend to be most adequate in relation to PALM and FOCAL. Since the main objective of the analysis is to clarify if FOCAL indeed improves on logit, I take the conservative position and adopt the model favoring logit (i.e. token-based utilities), but as indicated above and shown below, the results are robust in the sense that alternative assumptions favor FOCAL even more.

Table 13: Are utility scales adaptive? Accuracy in-sample (BIC, less is better), all experiments pooled

	Heterogenous pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
Logit	7651.6	$\approx$	7728	$\approx$	7735	7650.5	$\approx$	7710	$\approx$	7730.3
FOCAL	5830.6	$\approx$	5901.2	$\approx$	5839.2	5811.4	$\approx$	5849.6	$\approx$	5813.4
PALM	7445.7	$\approx$	7526.2	$\approx$	7516.1	7427.8	$\approx$	7466.3	$\approx$	7474.2
OGEV	7363.6	$\approx$	7432.9	$\approx$	7388.8	7333.8	$\approx$	7332.9	$\approx$	7334.6

Note: The table provides the goodness-of-fit in sample (BIC) for $2 \times 3 \times 4$ models. The extent of subject heterogeneity varies between heterogeneous preferences & precision, and full heterogeneity (preferences, precision and choice). The second dimension concerns the "utility standardization": token-based, value-based, and contextual utility. The third dimension is the set of models considered. The focus of this table is on utility standardization, for which the results of likelihood ratio tests following Schennach and Wilhelm (2016) are provided. Throughout, $\approx$ indicates p-values above 0.05, $>, <$ indicate p-values between 0.005 and 0.05, and $\gg, \ll$ indicate p-values below 0.005. The latter roughly implements the Bonferroni correction per panel.

Table 14: Are utility scales adaptive? In-sample and counterfactual reliability, all experiments pooled

(a) Logit

Objective	Heterog. pref & prec					Full heterogeneity				
	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	7651.6	≈	7728	≈	7735	7650.5	≈	7710	≈	7730.3
Reliability pref	11419.4	≈	11531.6	≈	11443.7	11409.2	≈	11506.9	≈	11422.6
Rel pref & prec	11597.4	≪	16378.8	≫	11594.8	11584.4	≪	16340.6	≫	11535.6
Rel all pars	4085.5	≈	4128.3	≈	4101.1	4053.9	≈	4124	≈	4100.6

(b) FOCAL

Objective	Heterog. pref & prec					Full heterogeneity				
	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	5830.6	≈	5901.2	≈	5839.2	5811.4	≈	5849.6	≈	5813.4
Reliability pref	6522	≈	6583.9	≈	6418	6493.7	≈	6584.4	≈	6436.4
Rel pref & prec	6561.3	≈	6621.3	≈	6501.7	6572.1	≈	6584.9	≈	6519.9
Rel all pars	1192.4	≈	1274.4	≈	1281.3	1180.7	<	1276.7	≈	1272.8

(c) PALM

Objective	Heterog. pref & prec					Full heterogeneity				
	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	7445.7	≈	7526.2	≈	7516.1	7427.8	≈	7466.3	≈	7474.2
Reliability pref	11340.6	≈	11457.2	≈	11162.7	11314.7	≈	11331	≈	11138.9
Rel pref & prec	11507.9	≈	11489.9	≈	11295.9	11507.8	≈	11452.9	≈	11253.4
Rel all pars	3820	≈	3880.4	≈	3860.1	3834.2	≈	3864.2	≈	3822.2

(d) OGEV

Objective	Heterog. pref & prec					Full heterogeneity				
	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	7363.6	≈	7432.9	≈	7388.8	7333.8	≈	7332.9	≈	7334.6
Reliability pref	10694.1	≈	11121.1	≈	10769.3	10524.6	≈	10681.5	≈	10604.5
Rel pref & prec	10882.6	≈	11214.9	≈	10924.4	10947.9	≈	10852.7	≈	10883
Rel all pars	3343.2	<	3445.2	≈	3340.3	3517.3	≈	3578.6	≈	3536.7

Note: This table extends Table 13 by additionally reporting on out-of-sample robustness of estimates: prediction of either preferences or preferences and precision across experiments (based on estimates from the "rich" experiments, as described above). The log-likelihoods are pooled across experiments. In the analysis of the reliability of all parameters, including the measure of the choice bias, the analysis focuses on the numerical DG experiments, i.e. the "graphical" DG experiment of FKM07 is neglected here (also as described above).

Table 15: Are utility scales adaptive? In-sample and counterfactual reliability, AM02

(a) Logit

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	1201.3	≈	1264.5	≈	1241.7	1200.4	≈	1262.6	≈	1241.2
Reliability pref	2106.1	≈	2139.8	≈	2141	2105.7	≈	2124.7	≈	2098
Rel pref & prec	2087.5	≪	2436.7	≫	2136.3	2087.4	≪	2340.6	≫	2099.6
Rel all pars	2270.1	≈	2306.8	≈	2309.6	2270.1	≈	2306.6	≈	2309.8

(b) FOCAL

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	321	≈	348.5	≈	339.6	317.8	≈	348	≈	318.5
Reliability pref	528.2	≈	564.8	≈	565.4	524.7	≈	557.1	≈	555.1
Rel pref & prec	521.9	<	604.4	≈	561	505	≪	591.2	≈	553.3
Rel all pars	599.7	≪	699.9	≈	651.3	582.7	≪	696.5	>	640.7

(c) PALM

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	1066.9	≈	1113.3	≈	1107.3	1065.1	≈	1095.9	≈	1082.8
Reliability pref	1929.5	≈	2010.3	≈	2005.9	1930.4	≈	1961.5	≈	1911.6
Rel pref & prec	1920.4	≈	1986.8	≈	1977.1	1919.1	≈	1961	≈	1926
Rel all pars	2095.7	≈	2140.8	≈	2154.7	2094.2	≈	2129	≈	2130.4

(d) OGEV

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	966.6	≈	984.1	≈	964	947.6	≈	954.4	≈	958.2
Reliability pref	1586.5	≪	1721.9	≈	1688.5	1758.2	≈	1788.3	≈	1749.4
Rel pref & prec	1667.1	<	1780.8	≈	1713.5	1790.2	≈	1814.7	≈	1778.2
Rel all pars	1641.5	<	1740.5	≈	1650.6	1841.2	≈	1843.7	≈	1849.8

Note: This table closely relates to Table 14, but focuses on the data of AM02. Note that “prediction of preferences” and “prediction of preferences and precision” refer to the accuracy of predicting the AM02 data set using estimates from the other rich dictator games (HJ06 and FKM07).

Table 16: Are utility scales adaptive? In-sample and counterfactual reliability, HJ06

(a) Logit

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	677.8	≈	675.7	≈	679	677.8	≈	675.7	≈	679
Reliability pref	861.7	≈	852.5	≈	863.2	861.7	≈	853.1	≈	861.2
Rel pref & prec	878.3	$<$	959.5	≈	875.8	877.4	≈	929.1	≈	873.9
Rel all pars	960.4	≈	965.6	≈	954.8	959.2	≈	955.8	≈	954.6

(b) FOCAL

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	246.2	≈	242.9	≈	241.5	242	≈	235	≈	237.5
Reliability pref	295.9	≈	288.5	≈	291.8	295.6	≈	284.7	≈	289.8
Rel pref & prec	292.6	≈	303.3	≈	298.9	292	≈	300.5	≈	295.7
Rel all pars	354.9	≈	364.4	≈	364.5	351.9	≈	364.6	≈	357.4

(c) PALM

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	630.9	≈	619	≈	628.6	615.4	≈	614.2	≈	616.7
Reliability pref	796.5	≈	795.5	≈	806.2	806	≈	789.1	≈	797.5
Rel pref & prec	821	≈	813.1	≈	809.6	820.1	≈	811.3	≈	808.5
Rel all pars	904.3	≈	894	≈	885.4	902.9	≈	896	≈	887.2

(d) OGEV

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	675.1	≈	673	≈	668.8	661.2	≈	661.1	≈	658.1
Reliability pref	797	≈	848.5	≈	786.8	788.8	≈	795	≈	795.9
Rel pref & prec	829.4	≈	867.7	≈	806.3	813.7	≈	817.5	≈	813.6
Rel all pars	898.9	≈	888.3	≈	878	872.6	≈	881.8	≈	866.3

Note: This table closely relates to Table 14, but focuses on the data of HJ06. Note that "prediction of preferences" and "prediction of preferences and precision" refer to the accuracy of predicting the HJ06 data set using estimates from the other rich dictator games (AM02 and FKM07).

Table 17: Are utility scales adaptive? In-sample and counterfactual reliability, FKM07

(a) Logit

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	5102.2	≈	5118.2	≈	5111.1	5102.1	≈	5102.1	≈	5107
Reliability pref	7612.9	≈	7664.1	≈	7556.3	7611.7	≈	7660.4	≈	7595.3
Rel pref & prec	7782.6	≪	12021.1	≫	7747.3	7790.9	≪	12127.5	≫	7734
Rel all pars	7774.3	≪	12012.8	≫	7739.1	7782.6	≪	12119.2	≫	7725.7

(b) FOCAL

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	5093.7	≈	5143.1	≈	5089.6	5082	≈	5099.9	≈	5089.1
Reliability pref	5498.4	≈	5526	≈	5366.2	5469.5	≈	5538.8	≈	5392.7
Rel pref & prec	5536.1	≈	5516.6	≈	5412.5	5559.1	≈	5498.6	≈	5434.3
Rel all pars	14082.7	≈	14016.8	≈	14518.1	12350.6	≈	13093.2	≈	11778

(c) PALM

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	5098.3	≈	5144.8	≈	5099	5097.6	≈	5107.1	≈	5093.5
Reliability pref	7790.7	≈	7791	≈	7513.2	7764	≈	7731.5	≈	7590.1
Rel pref & prec	7951.5	≈	7861.6	≈	7683.9	7942.2	≈	7850.3	≈	7719.8
Rel all pars	7915.9	≈	7813.9	≈	7666.5	7918.3	≈	7803	≈	7709.3

(d) OGEV

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	5080.5	≈	5138.9	≈	5084.3	5065.9	≈	5059.1	≈	5025.8
Reliability pref	7514	≈	7725.6	≈	7477.4	7195.6	≈	7232.3	≈	7226.3
Rel pref & prec	7614.3	≈	7797	≈	7622.6	7540.8	≈	7382.8	≈	7472.9
Rel all pars	7429.1	≈	7386.8	≈	7376.1	7366.9	≈	7435.1	≈	7330

Note: This table closely relates to Table 14, but focuses on the data of FKM07. Note that "prediction of preferences" and "prediction of preferences and precision" refer to the accuracy of predicting the FKM07 data set using estimates from the other rich dictator games (AM02 and HJ06).

Table 18: Are utility scales adaptive? In-sample and counterfactual reliability, CHST07

(a) Logit

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	670.3	≈	669.7	≈	703.1	670.3	≈	669.6	≈	703.1
Reliability pref	838.6	≈	875.2	≈	883.2	830.1	≈	868.7	≈	868.1
Rel pref & prec	848.9	<	961.5	>	835.4	828.6	<	943.5	>	828.1
Rel all pars	854.9	≈	855.9	≈	836.7	824.6	≈	861.6	≈	836.3

(b) FOCAL

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	169.7	≈	166.7	≈	168.5	169.6	≈	166.7	≈	168.4
Reliability pref	199.5	≈	204.5	≈	194.6	203.9	≈	203.8	≈	198.9
Rel pref & prec	210.7	≈	197	≈	229.3	216	≈	194.7	<	236.6
Rel all pars	237.8	≈	210.1	<	265.5	246	≈	215.5	<	274.7

(c) PALM

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	649.7	≈	649.1	≈	681.2	649.7	≈	649.1	≈	681.2
Reliability pref	824	≈	860.4	≈	837.5	814.3	≈	848.8	≈	839.8
Rel pref & prec	815	≈	828.4	≈	825.3	826.3	≈	830.2	≈	799.1
Rel all pars	820	≈	845.6	≈	820	837.1	≈	839.3	≈	804.7

(d) OGEV

	Heterog. pref & prec					Full heterogeneity				
Objective	Tokens		Contextual		Value	Tokens		Contextual		Value
In-sample BIC	641.4	≈	636.9	≈	671.7	659.1	≈	658.3	≈	692.4
Reliability pref	796.6	≈	825	≈	816.6	782	<	865.9	≈	832.9
Rel pref & prec	771.8	≈	769.5	≈	782.1	803.1	≈	837.6	≈	818.3
Rel all pars	802.8	≈	816.4	≈	811.7	803.5	≈	853.1	≈	820.6

Note: This table closely relates to Table 14, but focuses on the data of CHST07. Note that "prediction of preferences" and "prediction of preferences and precision" refer to the accuracy of predicting the CHST07 data set using estimates from the rich dictator games (AM02, HJ06 and FKM07).

5 In-sample accuracy of the choice models

The tables in this section provide all information on the basic "descriptive" accuracy of the various choice models. Table 19 reproduces the Table presented in the paper (for convenience), and Tables 20–22 provide the detailed information for each data set as well as the robustness checks for three models of subject heterogeneity and three models of utility standardization.

The main result on the descriptive adequacy is clear-cut and perfectly robust across robustness checks. In the graphical DG experiment, focal choice with respect to round numbers is not visible, which implies that all choice models are equally adequate. In the numerical DG experiments, where round number effects are pronounced, FOCAL generally fits highly significantly better than the other models, which in turn are almost equally adequate also in-sample but slightly favor PALM and OGEV over logit.

Table 19: Accuracy in capturing behavior in-sample (token-based utilities, full heterogeneity)

		Value range		Models						
		Clairvoyance	Random	FOCAL		OGEV		PALM		Logit
Large numerical	BIC	2812.1	8137	3371.9	$\ll$	4420.9	$\approx$	4492.6	$\ll$	4690.3
	R^2	1	0	0.895	$\gg$	0.698	$\approx$	0.684	$\gg$	0.647
Small numerical	BIC	260.8	1271.4	430.4	$\ll$	920	$\approx$	910.5	$\approx$	931.1
	R^2	1	0	0.832	$\gg$	0.348	$\approx$	0.357	$\approx$	0.337
Graphical	BIC	10021.8	23249.2	15103.9	$\approx$	15087.7	$\approx$	15119.4	$\approx$	15123.9
	R^2	1	0	0.616	$\approx$	0.617	$\approx$	0.615	$\approx$	0.614

Note: For each choice model (Logit, OGEV, PALM, and FOCAL), the descriptive accuracy (BIC and pseudo-R^2 in-sample) is reported for data from "Numerical DG" (AM02, HJ06, CHST07) and "Graphical DG" (FKM07). Significance of differences (between neighboring models) is evaluated by likelihood ratio tests following Schennach and Wilhelm (2016). Throughout, $\approx$ indicates p-values above 0.05, $>,<$ indicate p-values between 0.005 and 0.05, and $\gg,\ll$ indicate p-values below 0.005.

Table 20: Accuracy in capturing behavior in-sample: robustness checks (token-based utility)

(a) BIC, preference and precision heterogeneity

	LL-Range		Bayes Information Criterion (less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3042.4	≫	2162.2	≪	2807.8	<	2908	≪	3042.4
HJ06	970.9	2322.6	1648.8	≫	1217.1	≪	1646	≈	1601.8	≈	1648.8
FKM07	10021.8	23249.2	15124.1	≈	15115.5	≈	15102.3	≈	15120.1	≈	15124.1
CHST07	260.8	1271.4	931.1	≫	430.5	≪	902.3	≈	910.5	≈	931.1

(b) Pseudo-R^2, preference and precision heterogeneity

	LL-Range		Variance explained (more is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1	0	0.698	≪	0.919	≫	0.757	>	0.731	≫	0.698
HJ06	1	0	0.499	≪	0.818	≫	0.501	≈	0.533	≈	0.499
FKM07	1	0	0.614	≈	0.615	≈	0.616	≈	0.615	≈	0.614
CHST07	1	0	0.337	≪	0.832	≫	0.365	≈	0.357	≈	0.337

(c) BIC, preference heterogeneity

	LL-Range		Bayes Information Criterion (less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3595.8	≫	2243.7	≪	3151.1	≪	3444.5	≪	3595.8
HJ06	970.9	2322.6	1716.6	≫	1256.6	≪	1687.1	≈	1654.2	<	1716.6
FKM07	10021.8	23249.2	15710.9	≈	15697.6	≈	15673.2	≈	15709.4	≈	15710.9
CHST07	260.8	1271.4	1052.9	≫	440.5	≪	994	≈	1029.4	≈	1052.9

(d) Pseudo-R^2, preference heterogeneity

	LL-Range		Variance explained (more is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1	0	0.558	≪	0.899	≫	0.67	≫	0.596	≫	0.558
HJ06	1	0	0.448	≪	0.789	≫	0.47	≈	0.495	>	0.448
FKM07	1	0	0.57	≈	0.571	≈	0.573	≈	0.57	≈	0.57
CHST07	1	0	0.216	≪	0.822	≫	0.274	≈	0.239	≈	0.216

(e) BIC, full heterogeneity

	LL-Range		Bayes Information Criterion (less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3041.5	≫	2158.9	≪	2788.8	<	2906.3	≪	3041.5
HJ06	970.9	2322.6	1648.8	≫	1213	≪	1632.2	≈	1586.4	<	1648.8
FKM07	10021.8	23249.2	15123.9	≈	15103.9	≈	15087.7	≈	15119.4	≈	15123.9
CHST07	260.8	1271.4	931.1	≫	430.4	≪	920	≈	910.5	≈	931.1

(f) Pseudo-R^2, full heterogeneity

	LL-Range		Variance explained (more is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1	0	0.698	≪	0.92	≫	0.762	>	0.732	≫	0.698
HJ06	1	0	0.499	≪	0.821	≫	0.511	≈	0.545	>	0.499
FKM07	1	0	0.614	≈	0.616	≈	0.617	≈	0.615	≈	0.614
CHST07	1	0	0.337	≪	0.832	≫	0.348	≈	0.357	≈	0.337

Note: The layout is as in Table 19, but the results of the robustness checks are reported (alternative models of heterogeneity and token-based utility) and the results are broken down to each experiment for full clarity.

Table 21: Accuracy in capturing behavior in-sample: robustness checks (value-based utility)

(a) BIC, preference and precision heterogeneity

	LL-Range		Bayes Information Criterion (less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3082.9	$\gg$	2180.7	$\ll$	2805.2	$<$	2948.5	$<$	3082.9
HJ06	970.9	2322.6	1650	$\gg$	1212.4	$\ll$	1639.8	$\approx$	1599.5	$<$	1650
FKM07	10021.8	23249.2	15133	$\approx$	15111.5	$\approx$	15106.1	$\approx$	15120.9	$\approx$	15133
CHST07	260.8	1271.4	963.9	$\gg$	429.4	$\ll$	932.5	$\approx$	942	$\approx$	963.9

(b) Pseudo-R^2, preference and precision heterogeneity

	LL-Range		Variance explained (more is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1	0	0.687	$\ll$	0.915	$\gg$	0.757	$>$	0.721	$>$	0.687
HJ06	1	0	0.498	$\ll$	0.821	$\gg$	0.505	$\approx$	0.535	$>$	0.498
FKM07	1	0	0.614	$\approx$	0.615	$\approx$	0.616	$\approx$	0.615	$\approx$	0.614
CHST07	1	0	0.304	$\ll$	0.833	$\gg$	0.335	$\approx$	0.326	$\approx$	0.304

(c) BIC, preference heterogeneity

	LL-Range		Bayes Information Criterion (less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3551.4	$\gg$	2235.3	$\ll$	2969.7	$\ll$	3354.2	$\ll$	3551.4
HJ06	970.9	2322.6	1713.2	$\gg$	1252.4	$\ll$	1680	$\approx$	1650.3	$<$	1713.2
FKM07	10021.8	23249.2	15654.5	$\approx$	15621.5	$\approx$	15573.1	$\approx$	15642.8	$\approx$	15654.5
CHST07	260.8	1271.4	1064.8	$\gg$	438.6	$\ll$	1008.1	$\approx$	1040.2	$\approx$	1064.8

(d) Pseudo-R^2, preference heterogeneity

	LL-Range		Variance explained (more is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1	0	0.57	$\ll$	0.901	$\gg$	0.716	$\gg$	0.619	$\gg$	0.57
HJ06	1	0	0.451	$\ll$	0.792	$\gg$	0.475	$\approx$	0.497	$>$	0.451
FKM07	1	0	0.574	$\approx$	0.577	$\approx$	0.58	$\approx$	0.575	$\approx$	0.574
CHST07	1	0	0.204	$\ll$	0.824	$\gg$	0.261	$\approx$	0.229	$\approx$	0.204

(e) BIC, full heterogeneity

	LL-Range		Bayes Information Criterion (less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3082.3	$\gg$	2159.6	$\ll$	2799.4	$<$	2923.9	$\ll$	3082.3
HJ06	970.9	2322.6	1650	$\gg$	1208.4	$\ll$	1629.1	$\approx$	1587.7	$<$	1650
FKM07	10021.8	23249.2	15128.8	$\approx$	15111	$\approx$	15047.7	$\approx$	15115.3	$\approx$	15128.8
CHST07	260.8	1271.4	964	$\gg$	429.2	$\ll$	953.3	$\approx$	942	$\approx$	964

(f) Pseudo-R^2, full heterogeneity

	LL-Range		Variance explained (more is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1	0	0.688	$\ll$	0.92	$\gg$	0.759	$>$	0.727	$\gg$	0.688
HJ06	1	0	0.498	$\ll$	0.824	$\gg$	0.513	$\approx$	0.544	$>$	0.498
FKM07	1	0	0.614	$\approx$	0.615	$\approx$	0.62	$\approx$	0.615	$\approx$	0.614
CHST07	1	0	0.304	$\ll$	0.833	$\gg$	0.315	$\approx$	0.326	$\approx$	0.304

Note: The layout is as in Table 19, but the results of the robustness checks are reported (alternative models of heterogeneity and value-based utility) and the results are broken down to each experiment for full clarity.

Table 22: Accuracy in capturing behavior in-sample: robustness checks (contextual utility)

(a) BIC, preference and precision heterogeneity

	LL-Range		Bayes Information Criterion (less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3105.6	$\gg$	2189.6	$\ll$	2825.2	$<$	2954.5	$\ll$	3105.6
HJ06	970.9	2322.6	1646.7	$\gg$	1213.9	$\ll$	1643.9	$\approx$	1589.9	$<$	1646.7
FKM07	10021.8	23249.2	15140	$\approx$	15165	$\approx$	15160.7	$\approx$	15166.6	$\approx$	15140
CHST07	260.8	1271.4	930.5	$\gg$	427.5	$\ll$	897.7	$\approx$	909.9	$\approx$	930.5

(b) Pseudo-R^2, preference and precision heterogeneity

	LL-Range		Variance explained (more is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1	0	0.682	$\ll$	0.912	$\gg$	0.752	$>$	0.72	$\gg$	0.682
HJ06	1	0	0.5	$\ll$	0.82	$\gg$	0.502	$\approx$	0.542	$>$	0.5
FKM07	1	0	0.613	$\approx$	0.611	$\approx$	0.611	$\approx$	0.611	$\approx$	0.613
CHST07	1	0	0.337	$\ll$	0.835	$\gg$	0.37	$\approx$	0.358	$\approx$	0.337

(c) BIC, preference heterogeneity

	LL-Range		Bayes Information Criterion (less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3594.4	$\gg$	2286.9	$\ll$	3003	$\ll$	3432.5	$\ll$	3594.4
HJ06	970.9	2322.6	1707.3	$\gg$	1258.3	$\ll$	1681.4	$\approx$	1646.5	$<$	1707.3
FKM07	10021.8	23249.2	15711.2	$\approx$	15651.9	$\approx$	15651.4	$\approx$	15651.9	$\approx$	15711.2
CHST07	260.8	1271.4	1055	$\gg$	426.4	$\ll$	1009.1	$\approx$	1024.3	$\approx$	1055

(d) Pseudo-R^2, preference heterogeneity

	LL-Range		Variance explained (more is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1	0	0.559	$\ll$	0.888	$\gg$	0.708	$\gg$	0.599	$\gg$	0.559
HJ06	1	0	0.455	$\ll$	0.787	$\gg$	0.474	$\approx$	0.5	$>$	0.455
FKM07	1	0	0.57	$\approx$	0.574	$\approx$	0.574	$\approx$	0.574	$\approx$	0.57
CHST07	1	0	0.214	$\ll$	0.836	$\gg$	0.26	$\approx$	0.245	$\approx$	0.214

(e) BIC, full heterogeneity

	LL-Range		Bayes Information Criterion (less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3103.8	$\gg$	2189.1	$\ll$	2795.6	$\ll$	2937.1	$\ll$	3103.8
HJ06	970.9	2322.6	1646.7	$\gg$	1206	$\ll$	1632.1	$\approx$	1585.1	$<$	1646.7
FKM07	10021.8	23249.2	15123.9	$\approx$	15121.7	$\approx$	15080.9	$\approx$	15128.9	$\approx$	15123.9
CHST07	260.8	1271.4	930.5	$\gg$	427.5	$\ll$	919.1	$\approx$	909.9	$\approx$	930.5

(f) Pseudo-R^2, full heterogeneity

	LL-Range		Variance explained (more is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1	0	0.682	$\ll$	0.912	$\gg$	0.76	$\gg$	0.724	$\gg$	0.682
HJ06	1	0	0.5	$\ll$	0.826	$\gg$	0.511	$\approx$	0.546	$>$	0.5
FKM07	1	0	0.614	$\approx$	0.614	$\approx$	0.618	$\approx$	0.614	$\approx$	0.614
CHST07	1	0	0.337	$\ll$	0.835	$\gg$	0.349	$\approx$	0.358	$\approx$	0.337

Note: The layout is as in Table 19, but the results of the robustness checks are reported (alternative models of heterogeneity and contextual utility) and the results are broken down to each experiment for full clarity.

6 Plots and overview of parameter estimates

This section provides an overview of the parameter estimates for the main models (the full list of parameter estimates is provided at the end of the supplementary material) and plots illustrating the goodness-of-fit of the four choice models (Logit, OGEV, PALM, and FOCAL) in all treatments of all experiments.

Table 23 provides the mean degrees of altruism α and mean degrees of equity concerns β estimated under the four choice models and the three different assumptions of utility maximization. Since α is assumed to be normally distributed $\mathcal{N}(\mu_\alpha, \sigma_\alpha)$ but truncated at $[-0.5, 0, 5]$, the actual mean degree of altruism differs from the parameter μ_α reported below. The remaining tables in this section, Tables 24–26, provide the complete list of parameter estimates and associated standard errors for all choice models with full heterogeneity. The fitted choice distributions for the various choice models are provided in Figures 3–5 (for all experiments with well-defined treatments, i.e. for the numerical DG experiments AM02, HJ06, and CHST07).

Regarding the estimated mean degree of altruism (Table 27), the main result is that the range of estimates obtained across data sets is smallest for FOCAL, its range of estimates is substantially narrower than the ranges obtained under the other choice models. The mean degree of equity concerns is similar across models, and combined, these observations already suggest the estimates obtained under FOCAL are more reliable and predictive than the estimates obtained under the alternative models. The plots in Figures 3–5 may help clarify the reason: FOCAL indeed captures round-number effects, and in this sense it factors the "presentation effect" out of the choice propensities, while the other models do not capture the round-number effect and hence cannot factor it out.

Table 23: Mean degrees of altruism α and efficiency concerns β

(a) Full heterogeneity, token-based utility

	FOCAL		OGEV		PALM		Logit	
	α	β	α	β	α	β	α	β
AM02	0.08	0.17	-0.13	-0.16	-0.27	0.25	-0.07	-1.69
HJ06	0.23	-0.67	0.13	-0.6	0.17	-0.65	0.17	-0.65
FKM07	0.02	0.26	0.15	-0.17	0.01	0.24	0.01	0.24
CHST07	0.16	-0.14	0.5	-0.62	0.5	-0.62	0.5	-0.62

(b) Full heterogeneity, value-based utility

	FOCAL		OGEV		PALM		Logit	
	α	β	α	β	α	β	α	β
AM02	0.18	-0.02	-0.13	-0.23	-0.01	0.15	-0.02	-2.91
HJ06	0.19	-0.72	0.17	-0.72	0.18	-0.7	0.18	-0.7
FKM07	0.03	0.43	0.04	0.27	0.05	0.3	0.04	0.29
CHST07	0.16	-0.2	0.5	-1.22	0.5	-1.22	0.5	-1.22

(c) Full heterogeneity, contextual utility

	FOCAL		OGEV		PALM		Logit	
	α	β	α	β	α	β	α	β
AM02	0.12	0.27	-0.19	-0.12	-0.2	0.12	-0.19	-2.21
HJ06	0.2	-0.67	0.07	-0.66	0.2	-0.68	0.2	-0.68
FKM07	0.14	0.17	0.09	0.19	0.14	0.18	0.01	0.24
CHST07	0.15	-0.15	0.5	-0.62	0.5	-0.62	0.5	-0.62

Note: This table reports the estimated means of the degree of altruism (α) and the degree of equity concerns (β). The estimates are given for each choice model (Logit, OGEV, PALM, FOCAL) and each data set, allowing for all three forms of utility standardization. The estimated standard errors are mostly rather close to zero and skipped for readability of this table. The are reported in the tables toward the end of the supplementary material.

Table 24: All parameter estimates: Full heterogeneity, token-based utility

(a) Logit

	Altruism α		Efficiency β		Precision λ	
	μ_α	σ_α	μ_β	σ_β	μ_λ	σ_λ
AM02	−0.44 (0.002)	0.72 (0)	−1.69 (0.01)	32.14 (0.211)	0.35 (0.677)	4.91 (0.609)
HJ06	0.27 (0)	0.29 (0)	−0.65 (0)	6.48 (0.001)	−1.53 (0.115)	1.52 (0.056)
FKM07	0.07 (0)	0.88 (0)	0.24 (0)	1.66 (0)	−1.99 (0.034)	2.02 (0.016)
CHST07	0.69 (100)	0 (0)	−0.62 (0)	11.54 (0)	1.59 (0.832)	100 (0)

(b) OGEV

	Altruism α		Efficiency β		Precision λ		Correlation κ	
	μ_α	σ_α	μ_β	σ_β	μ_λ	σ_λ	μ_κ	σ_κ
AM02	−0.56 (0.021)	0.56 (0.021)	−0.16 (0.009)	85.8 (0)	−1.45 (0.119)	3.2 (0.172)	−3.71 (0.202)	0.04 (0.17)
HJ06	0.23 (0)	0.33 (0)	−0.6 (0.015)	6.18 (0)	−2.02 (0.234)	1.58 (0.134)	−0.05 (2.463)	3.73 (2.172)
FKM07	0.24 (0)	0.3 (0)	−0.17 (0)	1.67 (0)	−3.01 (0.05)	1.06 (0.014)	0 (0)	2.22 (0.009)
CHST07	0.73 (100)	0 (0)	−0.62 (0)	11.54 (0)	2.2 (2.843)	80.15 (33.612)	−4.61 (1.937)	0.11 (0.904)

(c) PALM

	Altruism α		Efficiency β		Precision λ		Perception κ	
	μ_α	σ_α	μ_β	σ_β	μ_λ	σ_λ	μ_κ	σ_κ
AM02	−0.46 (0)	0.27 (0)	0.25 (0.087)	77.51 (0.017)	−0.94 (0.148)	4.54 (1.418)	0.88 (100)	0.09 (100)
HJ06	0.27 (0.001)	0.29 (0)	−0.65 (0.003)	6.47 (0.014)	−1.54 (0.056)	1.51 (0.055)	0.43 (100)	0.11 (100)
FKM07	0.06 (0)	0.83 (0)	0.24 (0)	1.66 (0)	−2 (0.06)	2.05 (0.032)	−0.54 (0.005)	0.01 (0)
CHST07	0.69 (100)	0 (0)	−0.62 (0)	11.54 (0)	1.59 (1.021)	100 (0)	0.11 (100)	0.01 (100)

(d) FOCAL

	Altruism α		Efficiency β		Precision λ		Focality κ	
	μ_α	σ_α	μ_β	σ_β	μ_λ	σ_λ	μ_κ	σ_κ
AM02	0.16 (0.007)	0.37 (0.009)	0.17 (0.007)	7.35 (0.181)	−0.76 (0.121)	1.45 (0.096)	0.62 (0.036)	0.13 (0.056)
HJ06	0.41 (0)	0.3 (0)	−0.67 (0)	6.38 (0.001)	−1.44 (0.173)	1.52 (0.141)	0.42 (0.055)	0.21 (0.108)
FKM07	0.08 (0)	0.62 (0)	0.26 (0)	1.6 (0.002)	−1.93 (0.032)	1.88 (0.027)	−2.97 (0.08)	1.33 (0.018)
CHST07	1 (37.523)	0.67 (0)	−0.14 (0)	0.01 (0)	−2.66 (0.236)	1.28 (0.126)	0.82 (0.048)	0 (37.496)

Note: These tables report the full set of parameter estimates for each choice model. The estimates are given for each choice model and each data set, assuming full heterogeneity and token-based utility. For each "base parameter" $(\alpha, \beta, \lambda, \kappa)$, two numbers are given: μ and σ. These are the estimated means and standard deviations of the underlying normal distributions (as described above, α is truncated normal, β is normal, λ is log-normal, and κ is log-normal and censored at 1 in OGEV and PALM). The estimated Huber-Sandwich standard errors are provided in parentheses.

Table 25: All parameter estimates: Full heterogeneity, contextual utility

(a) Logit

	Altruism α		Efficiency β		Precision λ	
	μ_α	σ_α	μ_β	σ_β	μ_λ	σ_λ
AM02	−0.56	0.42	−2.21	36.13	4.07	3.25
	(0)	(0)	(0.003)	(0.025)	(0.19)	(0.24)
HJ06	0.43	0.35	−0.68	6.72	2.56	1.51
	(0)	(0)	(0)	(0.001)	(0.348)	(0.225)
FKM07	0.07	0.88	0.24	1.66	−1.98	2.02
	(0)	(0)	(0)	(0)	(0.027)	(0.039)
CHST07	0.69	0	−0.62	11.54	2.32	88.39
	(100)	(0)	(0)	(0)	(0.793)	(100)

(b) OGEV

	Altruism α		Efficiency β		Precision λ		Correlation κ	
	μ_α	σ_α	μ_β	σ_β	μ_λ	σ_λ	μ_κ	σ_κ
AM02	−0.69	0.49	−0.12	86.79	4.04	3.36	−3.8	0.37
	(0.168)	(0.108)	(1.109)	(100)	(4.146)	(4.629)	(1.687)	(3.074)
HJ06	0.23	0.49	−0.66	6.74	2.45	1.36	−0.26	3.81
	(0.001)	(0.001)	(0.003)	(0.013)	(0.116)	(0.086)	(0.672)	(0.981)
FKM07	0.17	0.34	0.19	1.02	3.69	1.33	−2.74	3.74
	(0)	(0)	(0)	(0)	(0.22)	(0.01)	(0.08)	(0.095)
CHST07	0.73	0	−0.62	11.54	3.32	85.26	−4.61	0.11
	(100)	(0)	(0)	(0)	(0.569)	(63.87)	(1.482)	(1.636)

(c) PALM

	Altruism α		Efficiency β		Precision λ		Perception κ	
	μ_α	σ_α	μ_β	σ_β	μ_λ	σ_λ	μ_κ	σ_κ
AM02	−0.27	0.23	0.12	68.75	4.8	2.66	2.28	0.08
	(0)	(0)	(0.011)	(0.002)	(0.209)	(0.162)	(100)	(100)
HJ06	0.43	0.34	−0.68	6.72	2.55	1.52	0.4	0.11
	(0)	(0)	(0)	(0.081)	(0.278)	(0.232)	(100)	(100)
FKM07	0.32	0.38	0.18	1.01	3.56	1.56	−3.71	0.49
	(0.001)	(0.001)	(0.001)	(0.002)	(0.06)	(0.073)	(0.444)	(0.091)
CHST07	0.69	0	−0.62	11.54	2.32	88.39	0.11	0.01
	(100)	(0)	(0)	(0)	(0.844)	(39.988)	(100)	(100)

(d) FOCAL

	Altruism α		Efficiency β		Precision λ		Focality κ	
	μ_α	σ_α	μ_β	σ_β	μ_λ	σ_λ	μ_κ	σ_κ
AM02	0.29	0.4	0.27	5.79	3.28	1.29	0.67	0.01
	(0)	(0)	(0.007)	(0)	(0.127)	(0.15)	(0.032)	(0.142)
HJ06	0.39	0.33	−0.67	6.66	2.76	1.31	0.41	0.22
	(0)	(0)	(0)	(0.001)	(0.107)	(0.141)	(0.064)	(0.102)
FKM07	0.32	0.39	0.17	1.02	3.59	1.6	−3.56	1.03
	(0.001)	(0.001)	(0)	(0)	(0.043)	(0.019)	(0.082)	(0.011)
CHST07	0.89	0.65	−0.15	0	3.37	0.87	0.8	0
	(0.203)	(0.08)	(0)	(0.001)	(0.343)	(0.285)	(0.045)	(0.006)

Note: These tables report the full set of parameter estimates for each choice model. The estimates are given for each choice model and each data set, assuming full heterogeneity and contextual utility. For each "base parameter" ($\alpha, \beta, \lambda, \kappa$), two numbers are given: μ and σ. These are the estimated means and standard deviations of the underlying normal distributions (as described above, α is truncated normal, β is normal, λ is log-normal, and κ is log-normal and censored at 1 in OGEV and PALM). The estimated Huber-Sandwich standard errors are provided in parentheses.

Table 26: All parameter estimates: Full heterogeneity, value-based utility

(a) Logit

	Altruism α		Efficiency β		Precision λ	
	μ_α	σ_α	μ_β	σ_β	μ_λ	σ_λ
AM02	−0.38 (0.019)	1.13 (0.057)	−2.91 (0.005)	34.87 (0.001)	1.62 (0.247)	5.55 (2.315)
HJ06	0.43 (0)	0.37 (0)	−0.7 (1.229)	6.26 (0)	0.67 (0.596)	1.56 (0.421)
FKM07	0.18 (0.002)	0.55 (0.004)	0.29 (0.002)	2.53 (0.002)	1.22 (0.319)	2.27 (0.228)
CHST07	0.52 (100)	0 (0)	−1.22 (0)	33.58 (0.001)	−0.07 (0.284)	98.85 (16.037)

(b) OGEV

	Altruism α		Efficiency β		Precision λ		Correlation κ	
	μ_α	σ_α	μ_β	σ_β	μ_λ	σ_λ	μ_κ	σ_κ
AM02	−0.47 (0)	0.5 (0)	−0.23 (0.019)	85.63 (0.076)	0.82 (0.098)	2.83 (0.158)	−3.7 (0.271)	0.13 (0.19)
HJ06	0.47 (0)	0.42 (0)	−0.72 (0)	6.34 (0.001)	0.33 (0.165)	1.6 (0.095)	−1.81 (0.507)	1.98 (0.503)
FKM07	0.28 (0)	0.71 (0)	0.27 (0)	2.44 (0)	1.12 (0.023)	2.47 (0.06)	−0.95 (0.012)	1.34 (0.016)
CHST07	0.6 (100)	0 (0)	−1.22 (88.177)	33.59 (100)	−0.17 (34.316)	88.98 (100)	−4.25 (6.04)	0.02 (6.273)

(c) PALM

	Altruism α		Efficiency β		Precision λ		Perception κ	
	μ_α	σ_α	μ_β	σ_β	μ_λ	σ_λ	μ_κ	σ_κ
AM02	−0.31 (1.734)	1.95 (5.716)	0.15 (0)	99.93 (0.03)	2.08 (1.118)	3.91 (2.827)	1.76 (100)	0.12 (100)
HJ06	0.43 (0)	0.37 (0)	−0.7 (0.825)	6.26 (0)	0.67 (0.484)	1.55 (0.148)	0.43 (100)	0.1 (100)
FKM07	0.2 (0)	0.58 (0)	0.3 (0)	2.55 (0.001)	1.18 (0.033)	2.26 (0.111)	−0.06 (0.004)	0.21 (0.007)
CHST07	0.52 (100)	0 (0)	−1.22 (0)	33.58 (0.001)	−0.06 (0.356)	99.03 (21.316)	0.09 (100)	0.01 (100)

(d) FOCAL

	Altruism α		Efficiency β		Precision λ		Focality κ	
	μ_α	σ_α	μ_β	σ_β	μ_λ	σ_λ	μ_κ	σ_κ
AM02	0.35 (0.055)	0.32 (0.03)	−0.02 (0.034)	3.74 (0.25)	1.91 (0.131)	1.1 (0.096)	0.66 (0.033)	0.29 (0.076)
HJ06	0.43 (0)	0.36 (0)	−0.72 (0)	6.49 (0.009)	0.7 (0.195)	1.57 (0.119)	0.42 (0.057)	0.19 (0.094)
FKM07	0.26 (0)	0.85 (0.001)	0.43 (0)	1.29 (0)	0.63 (0.011)	1.64 (0.017)	−2.22 (0.009)	0 (0)
CHST07	0.94 (0)	0.66 (0)	−0.2 (0)	0 (0)	−0.91 (0.271)	1.18 (0.138)	0.82 (0.048)	0 (0.007)

Note: These tables report the full set of parameter estimates for each choice model. The estimates are given for each choice model and each data set, assuming full heterogeneity and value-based utility. For each "base parameter" $(\alpha, \beta, \lambda, \kappa)$, two numbers are given: μ and σ. These are the estimated means and standard deviations of the underlying normal distributions (as described above, α is truncated normal, β is normal, λ is log-normal, and κ is log-normal and censored at 1 in OGEV and PALM). The estimated Huber-Sandwich standard errors are provided in parentheses.

Figure 3: Goodness of fit: AM02

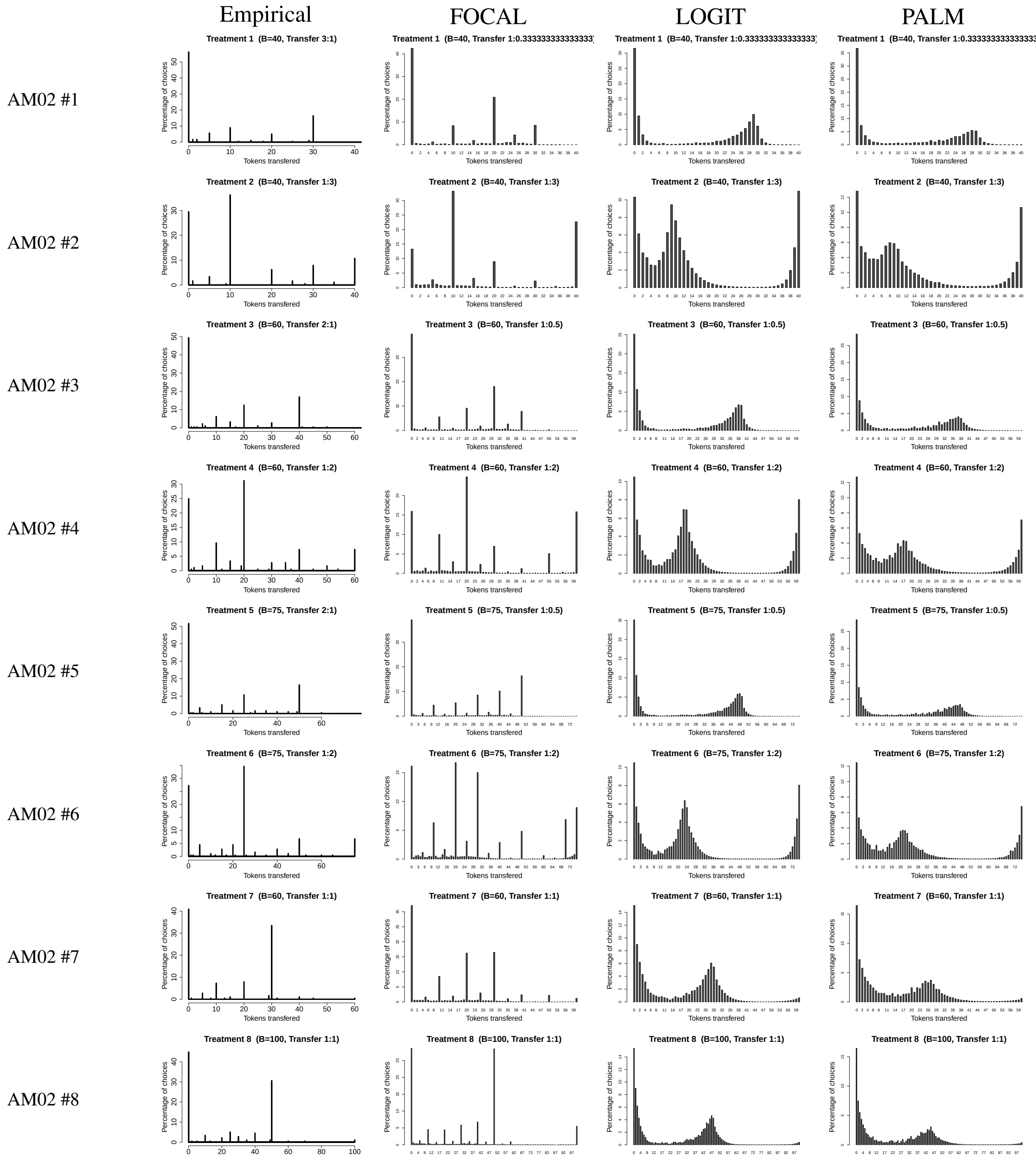


Figure 4: Goodness of fit: HJ06

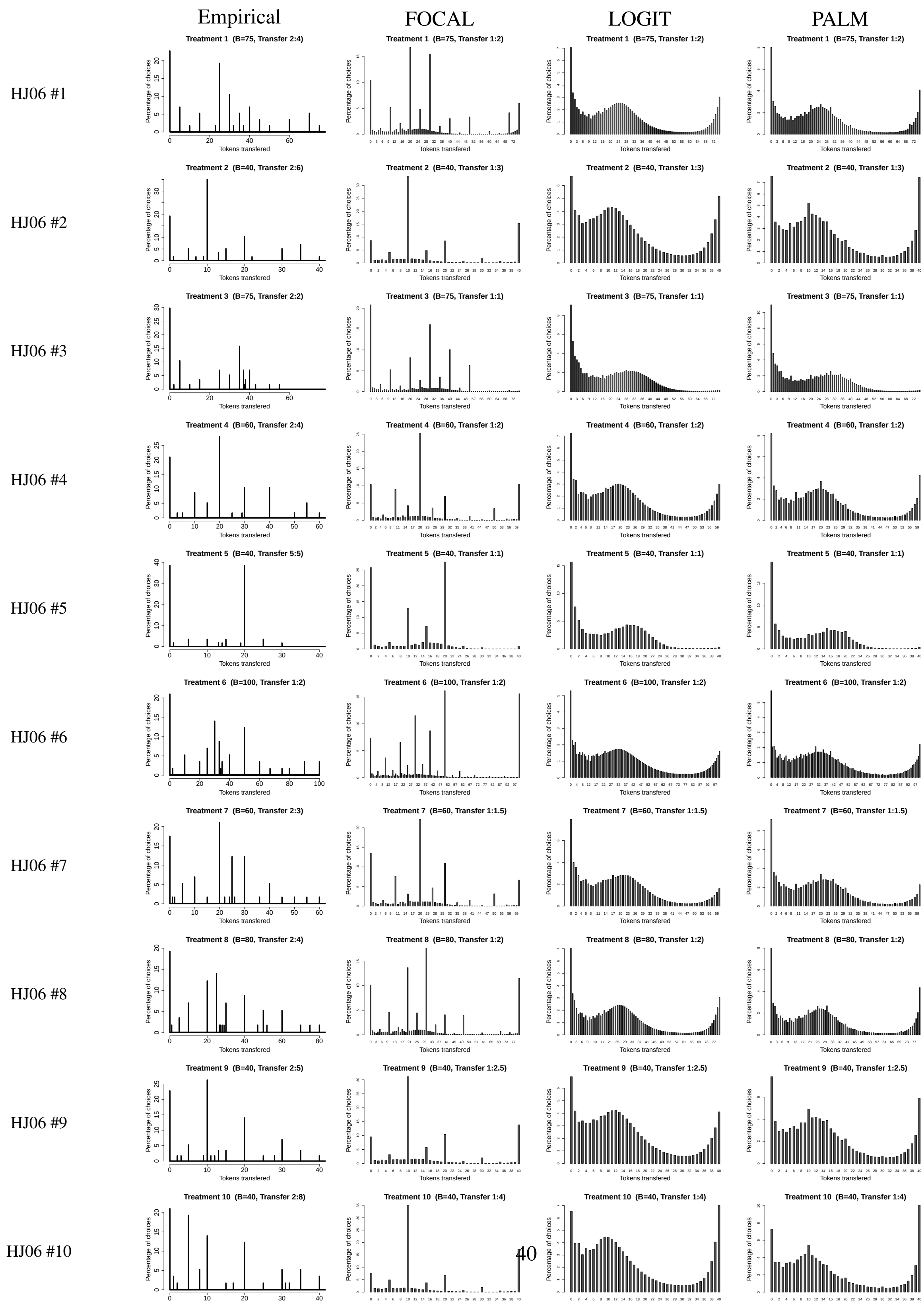
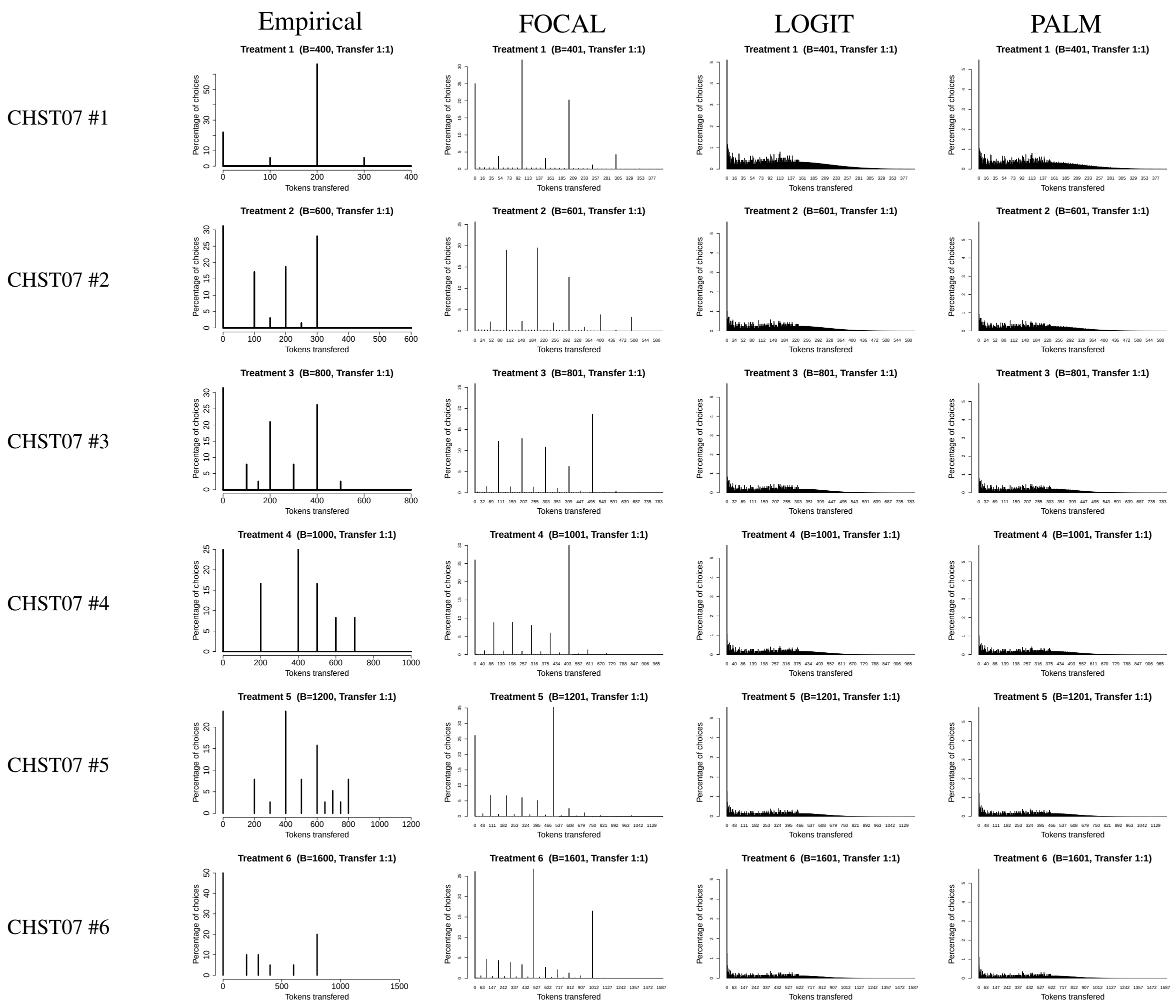


Figure 5: Goodness of fit: CHST07



7 Reliability of counterfactual predictions

This section provides additional information and robustness checks for the analysis of counterfactual reliability. Table 27 provides aggregate information including pseudo-R^2, while Tables 28–30 provide the detailed results for each data set as well as robustness checks for other assumptions on subject heterogeneity and utility standardization.

Section 1.5 defines counterfactual reliability formally. Briefly, the reliability of preference estimates is the BIC using the preference parameters out-of-sample and the remaining parameters (precision and choice) in-sample (the BIC penalty reflects the number of parameters used in-sample). As its name suggests, this focuses on the reliability of the preference estimates. The reliability of "preference and precision" estimates additionally uses the out-of-sample estimates of the precision estimates μ_λ and σ_λ for prediction. The accuracy of predictions when all parameters are used out-of-sample is evaluated as well.

The main result is that predictions based on estimates of FOCAL are generally highly significantly more reliable when predicting behavior out-of-sample than the estimates of the other models. This holds true for each data set, in aggregate, for all three degrees of counterfactual reliability, and in all robustness checks.

Table 27: Counterfactual reliability of estimates (token-based utilities, full heterogeneity)

(a) Preference estimates ($\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta$)

		Value range		Models						
		Clairvoyance	Random	FOCAL		OGEV		PALM		Logit
Large numerical	BIC	2812.1	8137	3632.4	$\ll$	5359.1	$\ll$	5548.6	$\ll$	5779.6
	R^2	1	0	0.846	$\gg$	0.522	$\gg$	0.486	$\gg$	0.443
Small numerical	BIC	260.8	1271.4	464.7	$\ll$	1042.8	$\approx$	1075.2	$\approx$	1090.9
	R^2	1	0	0.798	$\gg$	0.226	$\approx$	0.194	$\approx$	0.179
Graphical	BIC	10021.8	23249.2	15491.3	$\ll$	17217.5	$\approx$	17785.8	$\approx$	17633.5
	R^2	1	0	0.587	$\gg$	0.456	$\approx$	0.413	$\approx$	0.425

(b) Preference and precision estimates jointly ($\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta, \mu_\lambda, \sigma_\lambda$)

		Value range		Models						
		Clairvoyance	Random	FOCAL		OGEV		PALM		Logit
Large numerical	BIC	2812.1	8137	3609.1	$\ll$	5416	$\ll$	5551.4	$\ll$	5777
	R^2	1	0	0.85	$\gg$	0.511	$\gg$	0.486	$\gg$	0.443
Small numerical	BIC	260.8	1271.4	476.8	$\ll$	1064	$\approx$	1087.1	$\approx$	1089.5
	R^2	1	0	0.786	$\gg$	0.205	$\approx$	0.182	$\approx$	0.18
Graphical	BIC	10021.8	23249.2	15581	$\ll$	17562.7	$\approx$	17964.1	$\approx$	17812.7
	R^2	1	0	0.58	$\gg$	0.43	$\approx$	0.4	$\approx$	0.411

(c) All estimates jointly: Preference, precision, choice

		LL-Range		Bayes Information Criterion (less is better)						
		Clairvoyance	Random	FOCAL		OGEV		PALM		Logit
Large numerical	BIC	2812.1	8137	3746.8	$\ll$	5525.9	$\ll$	5809.2	$\ll$	6041.4
	R^2	1	0	0.824	$\gg$	0.49	$\gg$	0.437	$\gg$	0.394
Small numerical	BIC	260.8	1271.4	506.9	$\ll$	1064.3	$\approx$	1097.9	$\approx$	1085.4
	R^2	1	0	0.757	$\gg$	0.205	$\approx$	0.172	$\approx$	0.184

Note: For each choice model (Logit, OGEV, PALM, and FOCAL), the counterfactual reliability (BIC and pseudo-R^2 out-of-sample) of either the preference estimates, both preference and precision estimates, or all estimates are reported for data from "Numerical DG" (AM02, HJ06, CHST07) and "Graphical DG" (FKM07). Significance of differences (between neighboring models) is evaluated by likelihood ratio tests following Schennach and Wilhelm (2016). Throughout, $\approx$ indicates p-values above 0.05, $>, <$ indicate p-values between 0.005 and 0.05, and $\gg, \ll$ indicate p-values below 0.005.

Table 28: Counterfactual reliability of preference estimates (robustness checks)

(a) Preference and precision heterogeneity, value-based utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3982.2	$\gg$	2406.6	$\ll$	3529.7	$\ll$	3847	$<$	3982.2
HJ06	970.9	2322.6	1834.1	$\gg$	1262.7	$\ll$	1757.8	$\approx$	1777.1	$<$	1834.1
FKM07	10021.8	23249.2	17578.2	$\gg$	15388	$\ll$	17499.2	$\approx$	17535.1	$\approx$	17578.2
CHST07	260.8	1271.4	1144	$\gg$	455.4	$\ll$	1077.4	$\approx$	1098.3	$\approx$	1144

(b) Preference and precision heterogeneity, token-based utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3947.3	$\gg$	2369.3	$\ll$	3427.7	$\ll$	3770.6	$\ll$	3947.3
HJ06	970.9	2322.6	1832.7	$\gg$	1266.8	$\ll$	1767.9	$\approx$	1767.5	$<$	1832.7
FKM07	10021.8	23249.2	17634.7	$\gg$	15520.2	$\ll$	17535.9	$\approx$	17812.5	$\approx$	17634.7
CHST07	260.8	1271.4	1099.5	$\gg$	460.4	$\ll$	1057.4	$\approx$	1084.8	$\approx$	1099.5

(c) Preference and precision heterogeneity, contextual utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3981	$\gg$	2406	$\ll$	3563.1	$\ll$	3851.4	$<$	3981
HJ06	970.9	2322.6	1823.5	$\gg$	1259.5	$\ll$	1819.5	$\approx$	1766.5	$<$	1823.5
FKM07	10021.8	23249.2	17686	$\gg$	15547.9	$\ll$	17747.4	$\approx$	17812.9	$\approx$	17686
CHST07	260.8	1271.4	1136	$\gg$	465.3	$\ll$	1085.9	$\approx$	1121.2	$\approx$	1136

(d) Full heterogeneity, value-based utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3939.2	$\gg$	2396.2	$\ll$	3590.5	$\ll$	3752.8	$\ll$	3939.2
HJ06	970.9	2322.6	1832.1	$\gg$	1260.7	$\ll$	1766.9	$\approx$	1768.4	$<$	1832.1
FKM07	10021.8	23249.2	17617.2	$\gg$	15414.5	$\ll$	17248.2	$\approx$	17611.9	$\approx$	17617.2
CHST07	260.8	1271.4	1128.9	$\gg$	459.7	$\ll$	1093.7	$\approx$	1100.6	$\approx$	1128.9

(e) Full heterogeneity, token-based utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3946.9	$\gg$	2365.9	$\ll$	3599.4	$\ll$	3771.6	$\ll$	3946.9
HJ06	970.9	2322.6	1832.7	$\gg$	1266.5	$\ll$	1759.7	$\approx$	1777	$<$	1832.7
FKM07	10021.8	23249.2	17633.5	$\gg$	15491.3	$\ll$	17217.5	$\approx$	17785.8	$\approx$	17633.5
CHST07	260.8	1271.4	1090.9	$\gg$	464.7	$\ll$	1042.8	$\approx$	1075.2	$\approx$	1090.9

(f) Full heterogeneity, contextual utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3965.9	$\gg$	2398.3	$\ll$	3629.4	$\ll$	3802.6	$\ll$	3965.9
HJ06	970.9	2322.6	1824.1	$\gg$	1255.6	$\ll$	1765.9	$\approx$	1760.1	$<$	1824.1
FKM07	10021.8	23249.2	17682.2	$\gg$	15560.7	$\ll$	17254.2	$\approx$	17753.4	$\approx$	17682.2
CHST07	260.8	1271.4	1129.5	$\gg$	464.6	$\ll$	1126.8	$\approx$	1109.7	$\approx$	1129.5

Note: The layout is as in Table 19, but the results of the robustness checks are reported (alternative models of heterogeneity and utility standardization), focusing on the reliability of the preference estimates. The results are broken down to each experiment for full clarity.

Table 29: Counterfactual reliability of preference and precision estimates (robustness checks)

(a) Preference and precision heterogeneity, value-based utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3977.5	≫	2402.1	≪	3554.6	≪	3818.3	≪	3977.5
HJ06	970.9	2322.6	1846.7	≫	1269.9	≪	1777.2	≈	1780.5	<	1846.7
FKM07	10021.8	23249.2	17769.2	≫	15434.3	≪	17644.4	≈	17705.7	≈	17769.2
CHST07	260.8	1271.4	1096.2	≫	490.1	≪	1042.9	≈	1086.1	≈	1096.2

(b) Preference and precision heterogeneity, token-based utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3928.7	≫	2363.1	≪	3508.3	≪	3761.6	≪	3928.7
HJ06	970.9	2322.6	1849.3	≫	1263.6	≪	1800.3	≈	1791.9	<	1849.3
FKM07	10021.8	23249.2	17804.4	≫	15557.9	≪	17636.1	≈	17973.4	≈	17804.4
CHST07	260.8	1271.4	1109.7	≫	471.5	≪	1032.7	≈	1075.8	≈	1109.7

(c) Preference and precision heterogeneity, contextual utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	4277.9	≫	2445.5	≪	3621.9	≪	3827.9	≪	4277.9
HJ06	970.9	2322.6	1930.5	≫	1274.3	≪	1838.6	≈	1784.1	≪	1930.5
FKM07	10021.8	23249.2	22042.9	≫	15538.5	≪	17818.8	≈	17883.5	≪	22042.9
CHST07	260.8	1271.4	1222.4	≫	457.9	≪	1030.3	≈	1089.2	≪	1222.4

(d) Full heterogeneity, value-based utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3940.8	≫	2394.4	≪	3619.4	≪	3767.2	≪	3940.8
HJ06	970.9	2322.6	1844.9	≫	1266.6	≪	1784.5	≈	1779.5	<	1844.9
FKM07	10021.8	23249.2	17755.8	≫	15456.1	≪	17494.7	≈	17741.7	≈	17755.8
CHST07	260.8	1271.4	1089	≫	497.5	≪	1079.1	≈	1059.9	≈	1089

(e) Full heterogeneity, token-based utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	3928.6	≫	2346.2	≪	3631.4	<	3760.3	≪	3928.6
HJ06	970.9	2322.6	1848.4	≫	1262.9	≪	1784.7	≈	1791.1	<	1848.4
FKM07	10021.8	23249.2	17812.7	≫	15581	≪	17562.7	≈	17964.1	≈	17812.7
CHST07	260.8	1271.4	1089.5	≫	476.8	≪	1064	≈	1087.1	≈	1089.5

(f) Full heterogeneity, contextual utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	1841.2	5814.4	4181.8	≫	2432.4	≪	3655.9	≪	3802.2	≪	4181.8
HJ06	970.9	2322.6	1900	≫	1271.4	≪	1788.5	≈	1782.3	≪	1900
FKM07	10021.8	23249.2	22149.3	≫	15520.4	≪	17404.6	≈	17872.2	≪	22149.3
CHST07	260.8	1271.4	1204.3	≫	455.5	≪	1098.5	≈	1091.1	≪	1204.3

Note: The layout is as in Table 19, but the results of the robustness checks are reported (alternative models of heterogeneity and utility standardization), focusing on the reliability of the preference and precision estimates (jointly). The results are broken down to each experiment for full clarity.

Table 30: Counterfactual reliability of all estimates jointly (robustness checks)

(a) Preference and precision heterogeneity, value-based utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
Pooled	3072.9	9408.4	7174	$\gg$	4354.3	$\ll$	6413.3	$\ll$	6933	$\ll$	7174
AM02	1841.2	5814.4	4150.7	$\gg$	2492.5	$\ll$	3491.8	$\ll$	3995.8	$<$	4150.7
HJ06	970.9	2322.6	1925.8	$\gg$	1335.5	$\ll$	1849	$\approx$	1856.4	$<$	1925.8
CHST07	260.8	1271.4	1097.5	$\gg$	526.3	$\ll$	1072.5	$\approx$	1080.8	$\approx$	1097.5

(b) Preference and precision heterogeneity, token-based utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
Pooled	3072.9	9408.4	7158.4	$\gg$	4265.3	$\ll$	6416.2	$\ll$	6892.9	$\ll$	7158.4
AM02	1841.2	5814.4	4111.3	$\gg$	2440.9	$\ll$	3482.7	$\ll$	3936.8	$\ll$	4111.3
HJ06	970.9	2322.6	1931.4	$\gg$	1325.8	$\ll$	1869.8	$\approx$	1875.3	$<$	1931.4
CHST07	260.8	1271.4	1115.7	$\gg$	498.6	$\ll$	1063.6	$\approx$	1080.8	$\approx$	1115.7

(c) Preference and precision heterogeneity, contextual utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
Pooled	3072.9	9408.4	7201.3	$\gg$	4347.4	$\ll$	6518.2	$\ll$	6953.3	$\ll$	7201.3
AM02	1841.2	5814.4	4148	$\gg$	2541.1	$\ll$	3581.6	$\ll$	3982	$\ll$	4148
HJ06	970.9	2322.6	1936.6	$\gg$	1335.3	$\ll$	1859.3	$\approx$	1864.9	$<$	1936.6
CHST07	260.8	1271.4	1116.7	$\gg$	471	$\ll$	1077.3	$\approx$	1106.4	$\approx$	1116.7

(d) Full heterogeneity, value-based utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
Pooled	3072.9	9408.4	7173.6	$\gg$	4345.7	$\ll$	6609.7	$\ll$	6895.2	$\ll$	7173.6
AM02	1841.2	5814.4	4151	$\gg$	2481.9	$\ll$	3691	$\ll$	3971.6	$\ll$	4151
HJ06	970.9	2322.6	1925.5	$\gg$	1328.3	$\ll$	1837.2	$\approx$	1858.1	$<$	1925.5
CHST07	260.8	1271.4	1097.1	$\gg$	535.5	$\ll$	1081.4	$\approx$	1065.5	$\approx$	1097.1

(e) Full heterogeneity, token-based utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
Pooled	3072.9	9408.4	7126.8	$\gg$	4253.6	$\ll$	6590.2	$\ll$	6907.1	$\ll$	7126.8
AM02	1841.2	5814.4	4111.3	$\gg$	2423.9	$\ll$	3682.4	$\ll$	3935.3	$\ll$	4111.3
HJ06	970.9	2322.6	1930.1	$\gg$	1322.8	$\ll$	1843.5	$\approx$	1873.9	$<$	1930.1
CHST07	260.8	1271.4	1085.4	$\gg$	506.9	$\ll$	1064.3	$\approx$	1097.9	$\approx$	1085.4

(f) Full heterogeneity, contextual utility

	LL-Range		Counterfactual reliability (BIC, less is better)								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
Pooled	3072.9	9408.4	7197	$\gg$	4349.6	$\ll$	6651.6	$\ll$	6937.2	$\ll$	7197
AM02	1841.2	5814.4	4147.8	$\gg$	2537.7	$\ll$	3684.8	$\ll$	3970.1	$\ll$	4147.8
HJ06	970.9	2322.6	1926.8	$\gg$	1335.6	$\ll$	1852.8	$\approx$	1866.9	$<$	1926.8
CHST07	260.8	1271.4	1122.4	$\gg$	476.4	$\ll$	1114	$\approx$	1100.1	$\approx$	1122.4

Note: The layout is as in Table 19, but the results of the robustness checks are reported (alternative models of heterogeneity and utility standardization), focusing on the reliability of all estimates jointly. The results are broken down to each experiment for full clarity.

8 Consistency of estimates

This section mirrors the previous one but focuses on the consistency of the estimates. Table 31 provides aggregate results including pseudo-R^2, while Tables 32–34 provide the detailed results for each data set with robustness checks for other models of subject heterogeneity and utility standardization.

The difference to the previous section on reliability is that consistency refers the difference between out-of-sample BIC and in-sample BIC, and in this sense implements a likelihood ratio test of the significance of the difference between the in-sample estimates and the out-of-sample ones. Section 1.5 again provides formal definitions.

The main result confirms the above result on reliability, estimates of FOCAL are significantly more consistent than those of other models, in general, they are not significantly inconsistent. That is, they do not differ significantly from the in-sample estimates, i.e. the out-of-sample predictions are not significantly less accurate than the goodness-of-fits obtained in-sample. The exception are estimates from CHST07, which are weakly inconsistent in relation to the other estimates, i.e. the significance meets the 0.05 level but not the Bonferroni correction requiring roughly 0.005. In relation to the above result on counterfactual reliability, which show that CHST07 is predicted well using estimates from other experiments, the results in this section show that this reliability is less robust in the other direction. The former shows that subjects do not behave significantly differently across experiments (when we control for focality), while the latter shows that two observations per subject as in CHST07 do not suffice to reliably identify all parameters allowing for subject heterogeneity.

Table 31: Consistency of estimates (token-based utilities, full heterogeneity)

(a) Consistency of preference estimates (BIC)

		Value range		Models						
		Clairvoyance	Random	FOCAL		OGEV		PALM		Logit
Large numerical	BIC	0	22737.7	531.4	$\ll$	1601.6^{-}	$<$	2188.6^{--}	$>$	1655^{-}
	R^2	1	0	0.977	$\gg$	0.923^{-}	$>$	0.894^{--}	$<$	0.919^{-}
Small numerical	BIC	0	12910.4	1377.9^{-}	$\ll$	7111^{--}	$\ll$	8738.5^{--}	$\approx$	8890.6^{--}
	R^2	1	0	0.893^{-}	$\gg$	0.401^{--}	$\gg$	0.258^{--}	$\approx$	0.232^{--}
Graphical	BIC	0	5606	137.4	$\ll$	859.6^{--}	$>$	733.7^{--}	$\approx$	730.4^{--}
	R^2	1	0	0.975	$\gg$	0.789^{--}	$<$	0.817^{--}	$\approx$	0.807^{--}

(b) Consistency of preference and precision estimates (BIC)

		Value range		Models						
		Clairvoyance	Random	FOCAL		OGEV		PALM		Logit
Large numerical	BIC	0	22737.7	854.9	$\ll$	1825.5^{--}	$<$	2327.2^{--}	$>$	1755.7^{--}
	R^2	1	0	0.962	$\gg$	0.912^{--}	$>$	0.887^{--}	$<$	0.914^{--}
Small numerical	BIC	0	12910.4	1275.7^{-}	$\ll$	8108.3^{--}	$\ll$	9252.4^{--}	$\approx$	9359.2^{--}
	R^2	1	0	0.901^{-}	$\gg$	0.317^{--}	$\gg$	0.214^{--}	$\approx$	0.191^{--}
Graphical	BIC	0	5606	151.4	$\ll$	908.3^{--}	$\gg$	660.4^{--}	$\approx$	686.7^{--}
	R^2	1	0	0.973	$\gg$	0.777^{--}	$\ll$	0.835^{--}	$\approx$	0.819^{--}

(c) Consistency of all estimates jointly (BIC)

		LL-Range		Difference of BIC in- and out-of-sample (less is better)						
		Clairvoyance	Random	FOCAL		OGEV		PALM		Logit
Large numerical	BIC	0	6447	235.5^{-}	$\ll$	806.3^{--}	$\approx$	746^{--}	$\approx$	703.3^{--}
	R^2	1	0	0.963^{-}	$\gg$	0.818^{--}	$\approx$	0.829^{--}	$\approx$	0.83^{--}
Small numerical	BIC	0	4765.1	667^{--}	$\ll$	1692.2^{--}	$\ll$	2262^{--}	$\approx$	2307.5^{--}
	R^2	1	0	0.86^{--}	$\gg$	0.545^{--}	$\gg$	0.379^{--}	$\approx$	0.331^{--}

Note: For each choice model (Logit, OGEV, PALM, and FOCAL), the "consistency" of either the preference estimates or both preference and precision estimates are reported for data from "Numerical DG" (AM02, HJ06, CHST07) and "Graphical DG" (FKM07). The consistency is defined as the difference of BIC in-sample and the BIC using either preference or both preference and precision estimates out-of-sample (where the remaining parameters are still in-sample here, to focus on preference and precision). Significance of differences (between neighboring models) is evaluated by likelihood ratio tests following Schennach and Wilhelm (2016). Throughout, $\approx$ indicates p-values above 0.05, $>, <$ indicate p-values between 0.005 and 0.05, and $\gg, \ll$ indicate p-values below 0.005. In addition, the significance of the model's likelihood to that with the in-sample estimates of preference parameters is evaluated in likelihood ratio tests; significance at 0.005 is indicated by $^{--}$ and significance at 0.05 is indicated by $^{-}$.

Table 32: Consistency of preference estimates (robustness checks)

(a) Preference and precision heterogeneity, value-based utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	0	10090	720.7	≫	117.6	≪	841.5	≈	801.2	≈	720.7
HJ06	0	12613.5	754.5	≫	126.8	≪	898.3	>	680.3	≈	754.5
FKM07	0	5585.9	856.5^{--}	≫	159.7	≪	1061.6^{--}	≫	792.9^{--}	<	856.5^{--}
CHST07	0	12881.6	8794.6^{--}	≫	1332.4^{-}	≪	7340^{--}	≪	8665.5^{--}	≈	8794.6^{--}

(b) Preference and precision heterogeneity, token-based utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	0	10080.1	850.4	≫	268.8	≪	968.6	<	1427.3^{-}	>	850.4
HJ06	0	12626.8	816.9	≫	260.7	≪	713	≈	803.7	≈	816.9
FKM07	0	5598.6	746.9^{--}	≫	140	≪	1069.8^{--}	≫	728.3^{--}	≈	746.9^{--}
CHST07	0	12891.4	8889.1^{--}	≫	1404.8^{-}	≪	7240.1^{--}	≪	8725.4^{--}	≈	8889.1^{--}

(c) Preference and precision heterogeneity, contextual utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	0	10036.8	1026	≫	329.2	≪	1260.1^{-}	≈	1482.8^{-}	>	1026
HJ06	0	12552.9	846	≫	195.5	≪	956.3	≈	883.9	≈	846
FKM07	0	5577.4	738.7^{--}	≫	152.5	≪	1088.4^{--}	≫	805.3^{--}	≈	738.7^{--}
CHST07	0	12817.7	8800.2^{--}	≫	1370.8^{-}	≪	7759.8^{--}	≪	8621.1^{--}	≈	8800.2^{--}

(d) Full heterogeneity, value-based utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	0	10094.6	839.3	≫	135.2	≪	924.3	≈	956.3	≈	839.3
HJ06	0	12635.2	742	≫	213.9	≪	808.3	≈	682.6	<	742
FKM07	0	5611.1	685.4^{--}	≫	173^{-}	≪	863.9^{--}	≫	656.7^{--}	≈	685.4^{--}
CHST07	0	12907.2	8810.2^{--}	≫	1346.8^{-}	≪	7213.2^{--}	≪	8698.5^{--}	≈	8810.2^{--}

(e) Full heterogeneity, token-based utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	0	10095.9	839.3	≫	203.7	≪	902.7	<	1450.9^{-}	≫	839.3
HJ06	0	12641.8	815.6	≫	327.7	≪	698.9	≈	737.7	≈	815.6
FKM07	0	5606	730.4^{--}	≫	137.4	≪	859.6^{--}	>	733.7^{--}	≈	730.4^{--}
CHST07	0	12910.4	8890.6^{--}	≫	1377.9^{-}	≪	7111^{--}	≪	8738.5^{--}	≈	8890.6^{--}

(f) Full heterogeneity, contextual utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	0	10088	1053.9^{-}	≫	335.2	≪	1421.1^{-}	≈	1388.1^{-}	>	1053.9^{-}
HJ06	0	12596.6	789.1	≫	317	≪	995.3	≈	817.9	≈	789.1
FKM07	0	5585.7	797^{--}	≫	111.6	≪	1027^{--}	≫	752.6^{--}	≈	797^{--}
CHST07	0	12869.4	8750.6^{--}	≫	1440.8^{-}	≪	6602.4^{--}	≪	8635.5^{--}	≈	8750.6^{--}

Note: The layout is as in Table 31, but the results of the robustness checks are reported (alternative models of heterogeneity and utility standardization), focusing on the consistency of the preference estimates. The results are broken down to each experiment for full clarity.

Table 33: Consistency of joint preference and precision estimates (robustness checks)

(a) Preference and precision heterogeneity, value-based utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	0	10090	870.5	≫	301.7	≪	869.4	≈	924.7	≈	870.5
HJ06	0	12613.5	720.8	≫	162.6	≪	929	≫	607	<	720.8
FKM07	0	5585.9	678.7^{--}	≫	190.6^{-}	≪	915.1^{--}	≫	661^{--}	≈	678.7^{--}
CHST07	0	12881.6	9309.5^{--}	≫	1332.6^{-}	≪	7893.1^{--}	≪	9146.5^{--}	≈	9309.5^{--}

(b) Preference and precision heterogeneity, token-based utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	0	10080.1	941.7	≫	466.8	≪	991.1	<	1518.3^{-}	>	941.7
HJ06	0	12626.8	873	≫	300.3	≪	698.2	≈	789.6	≈	873
FKM07	0	5598.6	686.4^{--}	≫	159.5	≪	948.6^{--}	≫	644.2^{--}	≈	686.4^{--}
CHST07	0	12891.4	9336.2^{--}	≫	1265.5^{-}	≪	7919.1^{--}	≪	9234.6^{--}	≈	9336.2^{--}

(c) Preference and precision heterogeneity, contextual utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	0	10036.8	1745.3^{--}	≫	316.7	≪	1222.7^{-}	<	1508.8^{-}	≈	1745.3^{--}
HJ06	0	12552.9	12847.2^{--}	≫	182.4	≪	751.6	≈	658.3	≪	12847.2^{--}
FKM07	0	5577.4	2169.2^{--}	≫	150.2	≪	1064.3^{--}	≫	705.1^{--}	≪	2169.2^{--}
CHST07	0	12817.7	9190.7^{--}	≫	1511.1^{--}	≪	8307.4^{--}	<	9019.1^{--}	≈	9190.7^{--}

(d) Full heterogeneity, value-based utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	0	10094.6	869.6	≫	331.5	≪	912.7	≈	974.6	≈	869.6
HJ06	0	12635.2	725.5	≫	247.8	≪	721.3	≈	668.1	<	725.5
FKM07	0	5611.1	542.7^{--}	≫	205.6^{-}	≪	811.7^{--}	≫	537.9^{--}	≈	542.7^{--}
CHST07	0	12907.2	9278.2^{--}	≫	1334.2^{-}	≪	8199.4^{--}	≪	9157^{--}	≈	9278.2^{--}

(e) Full heterogeneity, token-based utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	0	10095.9	881.6	≫	497.1	≪	1074.8^{-}	<	1526.7^{-}	>	881.6
HJ06	0	12641.8	874.1	≫	357.8	≪	750.8	≈	800.6	<	874.1
FKM07	0	5606	686.7^{--}	≫	151.4	≪	908.3^{--}	≫	660.4^{--}	≈	686.7^{--}
CHST07	0	12910.4	9359.2^{--}	≫	1275.7^{-}	≪	8108.3^{--}	≪	9252.4^{--}	≈	9359.2^{--}

(f) Full heterogeneity, contextual utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
AM02	0	10088	2088.9^{--}	≫	318	≪	1361.1^{-}	≈	1449.5^{-}	≪	2088.9^{--}
HJ06	0	12596.6	12867.4^{--}	≫	240.4	≪	897.3	≈	750.7	≪	12867.4^{--}
FKM07	0	5585.7	1745.9^{--}	≫	98.6	≪	971.3^{--}	≫	682.4^{--}	≪	1745.9^{--}
CHST07	0	12869.4	9189.8^{--}	≫	1549.2^{--}	≪	7329.6^{--}	≪	9077.3^{--}	≈	9189.8^{--}

Note: The layout is as in Table 31, but the results of the robustness checks are reported (alternative models of heterogeneity and utility standardization), focusing on the consistency of the preference and precision estimates (jointly). The results are broken down to each experiment for full clarity.

Table 34: Consistency of all estimates jointly (robustness checks)

(a) Preference and precision heterogeneity, value-based utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
Pooled	0	11171.7	2954.6^{--}	≫	1063.5^{--}	≪	2071.6^{--}	≪	2885.9^{--}	≈	2954.6^{--}
AM02	0	1952.2	160.3	≈	152.7^{-}	≈	216.1^{-}	≈	175.6	≈	160.3
HJ06	0	4475.7	515.4^{--}	≫	124.7^{-}	≪	698.1^{--}	≫	494.7^{--}	≈	515.4^{--}
CHST07	0	4743.8	2278.9^{--}	≫	786.1^{--}	$<$	1157.5^{--}	≪	2215.6^{--}	≈	2278.9^{--}

(b) Preference and precision heterogeneity, token-based utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
Pooled	0	11197.1	3072.1^{--}	≫	911^{--}	≪	2120.2^{--}	≪	2945.1^{--}	$<$	3072.1^{--}
AM02	0	1946.4	220.3^{-}	$>$	108.6	$<$	222.6^{-}	≈	251^{-}	≈	220.3^{-}
HJ06	0	4493.1	544^{--}	≫	141.9^{-}	≪	577.2^{--}	$>$	448.1^{--}	≪	544^{--}
CHST07	0	4757.7	2307.9^{--}	≫	660.5^{--}	≪	1320.4^{--}	≪	2246^{--}	≈	2307.9^{--}

(c) Preference and precision heterogeneity, contextual utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
Pooled	0	11154.7	3037^{--}	≫	1032.7^{--}	≪	2302.5^{--}	≪	2997.9^{--}	≈	3037^{--}
AM02	0	1952.6	231.7^{-}	≫	59.2	≪	220^{-}	≈	250.4^{-}	≈	231.7^{-}
HJ06	0	4468.6	576.5^{--}	≫	120.6	≪	830.3^{--}	≫	555.2^{--}	≈	576.5^{--}
CHST07	0	4733.5	2228.8^{--}	≫	852.9^{--}	$<$	1252.2^{--}	≪	2192.4^{--}	≈	2228.8^{--}

(d) Full heterogeneity, value-based utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
Pooled	0	11222.2	2954.6^{--}	≫	1096.9^{--}	≪	2455.9^{--}	≪	2883^{--}	≈	2954.6^{--}
AM02	0	1956.3	159.2	≈	161.9^{-}	≈	188.7	≈	155.8	≈	159.2
HJ06	0	4496.9	515.9^{--}	≫	136.5^{-}	≪	593.5^{--}	$>$	475.8^{--}	$<$	515.9^{--}
CHST07	0	4768.9	2279.6^{--}	≫	798.6^{--}	≪	1673.7^{--}	≪	2251.4^{--}	≈	2279.6^{--}

(e) Full heterogeneity, token-based utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
Pooled	0	11212.1	3010.8^{--}	≫	902.6^{--}	≪	2498.6^{--}	≪	3007.9^{--}	≈	3010.8^{--}
AM02	0	1950.6	158.4	≈	117	≈	195^{-}	≪	265.4^{-}	≫	158.4
HJ06	0	4496.4	544.9^{--}	≫	118.5^{-}	≪	611.3^{--}	$>$	480.6^{--}	≪	544.9^{--}
CHST07	0	4765.1	2307.5^{--}	≫	667^{--}	≪	1692.2^{--}	≪	2262^{--}	≈	2307.5^{--}

(f) Full heterogeneity, contextual utility

	LL-Range		Likelihood-ratio: Difference in BIC in-sample and out-of-sample								
	Clairvoyance	Random	Logit		FOCAL		OGEV		PALM		Logit
Pooled	0	11171.5	3032.2^{--}	≫	1053.9^{--}	≪	2609.6^{--}	≪	3010.1^{--}	≈	3032.2^{--}
AM02	0	1960.5	224.5^{-}	≫	68.3	≪	247.9^{-}	≈	258^{-}	≈	224.5^{-}
HJ06	0	4469.1	578.4^{--}	≫	127.3^{-}	≪	706.2^{--}	≫	537.6^{--}	$<$	578.4^{--}
CHST07	0	4741.9	2229.3^{--}	≫	858.4^{--}	≪	1655.5^{--}	≪	2214.5^{--}	≈	2229.3^{--}

Note: The layout is as in Table 19, but the results of the robustness checks are reported (alternative models of heterogeneity and utility standardization), focusing on the reliability of all estimates jointly. The results are broken down to each experiment for full clarity.

9 Value of increasing model complexity

Tables 35–37 describe the evolution of BIC and Pseudo-R^2 as the model complexity increases, either as heterogeneity is extended or if logit is replaced by FOCAL. The analysis focuses on numerical DG experiments, as focality is insignificant in the GUI experiment. The tables provide the results for all three forms of utility standardization and all four degrees of predictive reliability: fully in-sample, preference prediction, preference and precision prediction, and full prediction. The results are essentially the same in all cases: For all forms of heterogeneity and extents of prediction, substituting FOCAL for logit reliably adds 30-40 percentage points to the pseudo-R^2, i.e. controlling for focality is about as critical as controlling for heterogeneity.

Table 35: Incremental contributions of the model components (token-based utility)

(a) Accuracy in-sample (BIC, Pseudo-R^2)

	Clairvoyance	Random	Homogenous		Het Prefs		Het Pref & Prec		Full Het
Logit	3558.9	9408.4	8166.6	$\gg$	6365.3	$\gg$	5622.3	$\approx$	5621.4
FOCAL	3558.9	9408.4	5878.2	$\gg$	3940.8	$>$	3809.8	$\approx$	3802.3
Logit			0.212	$\ll$	0.48	$\ll$	0.598	$\approx$	0.598
FOCAL			0.603	$\ll$	0.863	$<$	0.884	$\approx$	0.885

(b) Counterfactual reliability of preference estimates (BIC, Pseudo-R^2)

	Clairvoyance	Random	Homogenous		Het Prefs		Het Pref & Prec		Full Het
Logit	3558.9	9408.4	9940.9	$\gg$	6546.9	$\ll$	6879.4	$\approx$	6870.5
FOCAL	3558.9	9408.4	6108.8	$\gg$	4733.2	$\gg$	4096.5	$\approx$	4097.1
Logit			-0.091	$\ll$	0.452	$\gg$	0.399	$\approx$	0.401
FOCAL			0.564	$\ll$	0.738	$\ll$	0.838	$\approx$	0.838

(c) Counterfactual reliability of preference and precision estimates jointly (BIC, Pseudo-R^2)

	Clairvoyance	Random	Homogenous		Het Prefs		Het Pref & Prec		Full Het
Logit	3558.9	9408.4	9689.8	$\gg$	6902.8	$\approx$	6887.8	$\approx$	6866.5
FOCAL	3558.9	9408.4	7886.5	$\gg$	4486.7	$\gg$	4098.2	$\approx$	4085.9
Logit			-0.048	$\ll$	0.395	$\approx$	0.398	$\approx$	0.401
FOCAL			0.26	$\ll$	0.777	$\ll$	0.838	$\approx$	0.84

(d) Counterfactual reliability of all estimates jointly (BIC, Pseudo-R^2)

	Clairvoyance	Random	Homogenous		Het Prefs		Het Pref & Prec		Full Het
Logit	3558.9	9408.4	8339973.9	$\gg$	6926.6	$\approx$	7158.4	$\approx$	7126.8
FOCAL	3558.9	9408.4	17991.6	$\gg$	4590.8	$\gg$	4265.3	$\approx$	4253.6
Logit			-1424.145	$\ll$	0.392	$\approx$	0.355	$\approx$	0.36
FOCAL			-1.467	$\ll$	0.76	$\ll$	0.812	$\approx$	0.814

Note: This table comprises four panels (accuracy in-sample, and the three frameworks for measuring counterfactual reliability). Each panel provides measures of adequacy for overall 4×2 models: four models of subject heterogeneity (starting with the "representative agent" model, i.e. homogeneity) and the two central models of choice in this paper. The measures of adequacy are BIC (less is better) in the upper half of each panel, and the implied pseudo-R^2 (more is better) in the lower half. The respective likelihood-ratio tests are as before.

Table 36: Incremental contributions of the model components (robustness check: contextual utility)

(a) Accuracy in-sample (BIC, Pseudo-R^2)

	Clairvoyance	Random	Homogenous		Het Prefs		Het Pref & Prec		Full Het
Logit	3558.9	9408.4	8166.6	$\gg$	6356.8	$\gg$	5682.8	$\approx$	5680.9
FOCAL	3558.9	9408.4	5878.2	$\gg$	3971.5	$>$	3831	$\approx$	3822.6
Logit			0.212	$\ll$	0.482	$\ll$	0.588	$\approx$	0.588
FOCAL			0.603	$\ll$	0.858	$<$	0.88	$\approx$	0.882

(b) Counterfactual reliability of preference estimates (BIC, Pseudo-R^2)

	Clairvoyance	Random	Homogenous		Het Prefs		Het Pref & Prec		Full Het
Logit	3558.9	9408.4	9940.9	$\gg$	6619.5	$\ll$	6940.5	$\approx$	6919.4
FOCAL	3558.9	9408.4	6108.8	$\gg$	4508.3	$\gg$	4130.8	$\approx$	4118.5
Logit			-0.091	$\ll$	0.44	$\gg$	0.39	$\approx$	0.393
FOCAL			0.564	$\ll$	0.773	$\ll$	0.833	$\approx$	0.835

(c) Counterfactual reliability of preference and precision estimates jointly (BIC, Pseudo-R^2)

	Clairvoyance	Random	Homogenous		Het Prefs		Het Pref & Prec		Full Het
Logit	3558.9	9408.4	9689.8	$\gg$	7647	$\approx$	7430.7	$\approx$	7286.1
FOCAL	3558.9	9408.4	7886.5	$\gg$	4620.1	$\gg$	4177.7	$\approx$	4159.3
Logit			-0.048	$\ll$	0.278	$\approx$	0.312	$\approx$	0.335
FOCAL			0.26	$\ll$	0.756	$\ll$	0.826	$\approx$	0.829

(d) Counterfactual reliability of all estimates jointly (BIC, Pseudo-R^2)

	Clairvoyance	Random	Homogenous		Het Prefs		Het Pref & Prec		Full Het
Logit	3558.9	9408.4	8339973.9	$\gg$	6637.7	$\ll$	7201.3	$\approx$	7197
FOCAL	3558.9	9408.4	17991.6	$\gg$	4985.1	$\gg$	4347.4	$\approx$	4349.6
Logit			-1424.145	$\ll$	0.437	$\gg$	0.348	$\approx$	0.349
FOCAL			-1.467	$\ll$	0.698	$\ll$	0.799	$\approx$	0.798

Note: Similar to Table 35, but the results are reported for contextual utility as robustness check.

Table 37: Incremental contributions of the model components (robustness check: value-based utility)

(a) Accuracy in-sample (BIC, Pseudo-R^2)

	Clairvoyance	Random	Homogenous		Het Prefs		Het Pref & Prec		Full Het
Logit	3558.9	9408.4	8166.6	$\gg$	6329.4	$\gg$	5696.8	$\approx$	5696.3
FOCAL	3558.9	9408.4	5878.2	$\gg$	3926.3	$>$	3822.5	$\approx$	3797.3
Logit			0.212	$\ll$	0.486	$\ll$	0.586	$\approx$	0.586
FOCAL			0.603	$\ll$	0.865	$<$	0.882	$\approx$	0.886

(b) Counterfactual reliability of preference estimates (BIC, Pseudo-R^2)

	Clairvoyance	Random	Homogenous		Het Prefs		Het Pref & Prec		Full Het
Logit	3558.9	9408.4	9940.9	$\gg$	6584.8	$\ll$	6960.3	$\approx$	6900.2
FOCAL	3558.9	9408.4	6108.8	$\gg$	4524.7	$\gg$	4124.8	$\approx$	4116.7
Logit			-0.091	$\ll$	0.446	$\gg$	0.386	$\approx$	0.396
FOCAL			0.564	$\ll$	0.771	$\ll$	0.834	$\approx$	0.835

(c) Counterfactual reliability of preference and precision estimates jointly (BIC, Pseudo-R^2)

	Clairvoyance	Random	Homogenous		Het Prefs		Het Pref & Prec		Full Het
Logit	3558.9	9408.4	9689.8	$\gg$	7103.3	$\approx$	6920.4	$\approx$	6874.6
FOCAL	3558.9	9408.4	7886.5	$\gg$	4447.1	$\gg$	4162.1	$\approx$	4158.5
Logit			-0.048	$\ll$	0.364	$\approx$	0.393	$\approx$	0.4
FOCAL			0.26	$\ll$	0.783	$\ll$	0.828	$\approx$	0.829

(d) Counterfactual reliability of all estimates jointly (BIC, Pseudo-R^2)

	Clairvoyance	Random	Homogenous		Het Prefs		Het Pref & Prec		Full Het
Logit	3558.9	9408.4	8339973.9	$\gg$	7296.8	$\approx$	7174	$\approx$	7173.6
FOCAL	3558.9	9408.4	17991.6	$\gg$	4754.4	$\gg$	4354.3	$\approx$	4345.7
Logit			-1424.145	$\ll$	0.333	$\approx$	0.353	$\approx$	0.353
FOCAL			-1.467	$\ll$	0.735	$\ll$	0.798	$\approx$	0.799

Note: Similar to Table 35, but the results are reported for value-based utility as robustness check.

10 Is entropy reliably captured?

This section analyzes the models' predictions in relation to the entropy observed in the experiments. Table 38 provides an overview for token-based utilities, Tables 39–41 list the detailed results for all data sets. The tables focus on data sets from numerical DGs, as entropy is not a concern in the graphical DGs (FKM07).

The main result is that FOCAL generally captures entropy (and in this sense round-number effects) significantly better than all other models, both in-sample and out-of-sample. Further, the entropy implied by FOCAL does not differ significantly from the empirical entropy, not even out-of-sample.

Table 38: Is entropy reliably captured (token-based utility)

(a) In sample

	Empirical	FOCAL		OGEV		PALM		Logit
Standard DG	2.1	2.53	$<$	3.27^{--}	$\approx$	3.29^{--}	$\approx$	3.32^{--}
Entitle DG	1.65	2.34	$<$	4.51^{--}	$\approx$	4.57^{--}	$\approx$	4.58^{--}

(b) Preference prediction

	Empirical	FOCAL		OGEV		PALM		Logit
Standard DG	2.1	2.52	$<$	3.46^{--}	$\approx$	3.27^{--}	$\approx$	3.3^{--}
Entitle DG	1.65	3.2^{-}	$<$	5.6^{--}	$\approx$	5.47^{--}	$\approx$	5.45^{--}

(c) Preference and precision prediction

	Empirical	FOCAL		OGEV		PALM		Logit
Standard DG	2.1	2.57	$\ll$	3.54^{--}	$\approx$	3.37^{--}	$\approx$	3.4^{--}
Entitle DG	1.65	3.06^{-}	$<$	5.22^{--}	$\approx$	5.05^{--}	$\approx$	4.99^{--}

(d) Full prediction

	Empirical	FOCAL		OGEV		PALM		Logit
Standard DG	2.1	2.54	$<$	3.22^{--}	$\approx$	3.3^{--}	$\approx$	3.3^{--}
Entitle DG	1.65	3.46^{-}	$\approx$	5.12^{--}	$\approx$	5.03^{--}	$\approx$	4.92^{--}

Note: This table comprises four panels (accuracy in-sample, and the three frameworks for measuring counterfactual reliability) and each panel comprises statistics for two groups of experiments: Standard (AM02, HJ06), and Entitle (CHST07). Each panel relates the empirical entropy (estimated entropy averaged across all treatments in the respective experiments) to the respective predictions of the four choice models. Significance of differences (between neighboring models) is evaluated by matched-pairs Wilcoxon tests. Throughout, $\approx$ indicates p-values above 0.05, $>, <$ indicate p-values between 0.005 and 0.05, and $\gg, \ll$ indicate p-values below 0.005. In addition, Minus signs indicate that the model's prediction differ significantly from the empirical estimate; significance at 0.005 is indicated by $^{--}$ and significance at 0.05 is indicated by $^{-}$.

Table 39: Is entropy reliably captured (separately for each data set)

(a) In sample

	Empirical	FOCAL		OGEV		PALM		Logit
AM02	1.78	2.41	$\approx$	2.94^{-}	$\approx$	3.13^{-}	$\approx$	3.11^{-}
HJ06	2.36	2.63	$<$	3.52^{-}	$\approx$	3.42^{-}	$\approx$	3.49^{-}
CHST07	1.65	2.34	$<$	4.51^{--}	$\approx$	4.57^{--}	$\approx$	4.58^{--}

(b) Preference prediction

	Empirical	FOCAL		OGEV		PALM		Logit
AM02	1.78	2.37	$<$	3.45^{-}	$\approx$	3.46^{-}	$\approx$	3.33^{-}
HJ06	2.36	2.64	$\approx$	3.46^{-}	$\approx$	3.12	$\approx$	3.27^{-}
CHST07	1.65	3.2^{-}	$<$	5.6^{--}	$\approx$	5.47^{--}	$\approx$	5.45^{--}

(c) Preference and precision prediction

	Empirical	FOCAL		OGEV		PALM		Logit
AM02	1.78	2.5	$\approx$	3.6^{--}	$\approx$	3.55^{--}	$\approx$	3.6^{--}
HJ06	2.36	2.63	$\approx$	3.49^{-}	$\approx$	3.22	$\approx$	3.23
CHST07	1.65	3.06^{-}	$<$	5.22^{--}	$\approx$	5.05^{--}	$\approx$	4.99^{--}

(d) Full prediction

	Empirical	FOCAL		OGEV		PALM		Logit
AM02	1.78	2.7^{-}	$\approx$	3.68^{--}	$\approx$	3.61^{--}	$\approx$	3.66^{--}
HJ06	2.36	2.41	$\approx$	2.86	$\approx$	3.05	$\approx$	3.02
CHST07	1.65	3.46^{-}	$\approx$	5.12^{--}	$\approx$	5.03^{--}	$\approx$	4.92^{--}

Note: Similar to Table 38, but the results are reported for each data set individually.

Table 40: Is entropy reliably captured (robustness check: contextual utility)

(a) In sample

	Empirical	FOCAL		OGEV		PALM		Logit
AM02	1.78	2.36	$\approx$	2.62	$\approx$	2.58	$\approx$	3.1^{-}
HJ06	2.36	2.68	$\approx$	3.55^{-}	$\approx$	3.58^{-}	$\approx$	3.63^{-}
CHST07	1.65	2.32	$<$	4.63^{--}	$\approx$	4.7^{--}	$\approx$	4.7^{--}

(b) Preference prediction

	Empirical	FOCAL		OGEV		PALM		Logit
AM02	1.78	2.31	$\approx$	3.12^{-}	$\approx$	3.11^{-}	$\approx$	3.32^{-}
HJ06	2.36	2.76	$\approx$	3.47^{-}	$\approx$	3.55^{-}	$\approx$	3.54^{-}
CHST07	1.65	3.07^{-}	$<$	5.59^{--}	$\approx$	5.64^{--}	$\approx$	5.52^{--}

(c) Preference and precision prediction

	Empirical	FOCAL		OGEV		PALM		Logit
AM02	1.78	2.35	$\approx$	3.07^{-}	$\approx$	3.45^{-}	$\approx$	3.88^{--}
HJ06	2.36	2.68	$\approx$	3.08	$\approx$	3.04	$\approx$	3.6^{-}
CHST07	1.65	3.01^{-}	$<$	5.55^{--}	$\approx$	5.69^{--}	$\approx$	6.12^{--}

(d) Full prediction

	Empirical	FOCAL		OGEV		PALM		Logit
AM02	1.78	2.66	$\approx$	3.56^{-}	$\approx$	3.6^{--}	$\approx$	3.65^{--}
HJ06	2.36	2.41	$\approx$	2.76	$\approx$	2.72	$\approx$	3.12
CHST07	1.65	3.77^{-}	$\approx$	5.77^{--}	$\approx$	5.66^{--}	$\approx$	5.81^{--}

Note: Similar to Table 38, but the results are reported for each data set individually and for contextual utility (as robustness check).

Table 41: Is entropy reliably captured (robustness check: value-based utility)

(a) In sample

	Empirical	FOCAL		OGEV		PALM		Logit
AM02	1.78	2.3	$\approx$	2.86^{-}	$\approx$	2.98^{-}	$\approx$	3.35^{-}
HJ06	2.36	2.7	$\approx$	3.45^{-}	$\approx$	3.48^{-}	$\approx$	3.54^{-}
CHST07	1.65	2.33	$<$	4.57^{--}	$\approx$	4.6^{--}	$\approx$	4.61^{--}

(b) Preference prediction

	Empirical	FOCAL		OGEV		PALM		Logit
AM02	1.78	2.07	$<$	3.26^{-}	$\approx$	3.31^{-}	$\approx$	3.49^{-}
HJ06	2.36	2.57	$\approx$	3.17	$\approx$	3.24	$\approx$	3.36^{-}
CHST07	1.65	3.09^{-}	$<$	5.47^{--}	$\approx$	5.53^{--}	$\approx$	5.49^{--}

(c) Preference and precision prediction

	Empirical	FOCAL		OGEV		PALM		Logit
AM02	1.78	2.34	$\approx$	3.24^{-}	$\approx$	3.47^{--}	$\approx$	3.53^{--}
HJ06	2.36	2.46	$\approx$	3.06	$\approx$	3.06	$\approx$	3.28^{-}
CHST07	1.65	2.85^{-}	$<$	4.55^{--}	$\approx$	4.8^{--}	$\approx$	4.89^{--}

(d) Full prediction

	Empirical	FOCAL		OGEV		PALM		Logit
AM02	1.78	2.76^{-}	$\approx$	3.61^{--}	$\approx$	3.65^{--}	$\approx$	3.7^{--}
HJ06	2.36	2.34	$\approx$	2.7	$\approx$	2.88	$\approx$	3.26
CHST07	1.65	3.32^{-}	$\approx$	4.73^{--}	$\approx$	4.87^{--}	$\approx$	4.99^{--}

Note: Similar to Table 38, but the results are reported for each data set individually and for value-based utility (as robustness check).

11 Detailed list of parameter estimates

For each model and each data set, the following tables provide the full set of parameter estimates, including Huber Sandwich estimates of the associated standard errors and the absolute values of the models' log-likelihoods. Table 42 provides the estimates for the logit models, Table 43 provides the estimates for the FOCAL models, Table 45 provides the estimates for the Ordered GEV models, Table 44 provides the estimates for the PALM models. Each table is divided into four panels, one per data set. Each panel contains twelve rows, one for each of the 4×3 models estimated, i.e. for each combination of the four specifications of heterogeneity (just preferences are heterogeneous; preferences and precision; preference, precision and choice; preference and choice – abbreviated below as Mix1, Mix2, Mix3, and Mix4, respectively) with the three specifications of utility standardization (token-based, value-based, contextual). The respective models are defined in the first column of all tables, which uniquely identifies them All choice models are formally defined in Section 1.1, the utility standardizations are specified in Section 1.2, and the forms of heterogeneity are specified in Section 1.3. In all cases, only the cells for the parameters actually used in the analysis contain entries, the remaining cells are empty. This, too, uniquely identifies the model estimated in a given row.

Table 42: Estimated parameters of **logit** models

	λ	κ	τ	α	β	σ_α	σ_β	σ_λ	$\lvert LL \rvert$
Dictator games I (AM02), Logit									
Mix1, Token:	0.383 (24.696)			−0.468 (334.226)	−1.056 (342.62)	4.467 (1744.839)	9.139 (0.063)		3570.38
Mix2, Token:	0.332 (0.652)			−0.436 (0)	−1.68 (0)	0.721 (0)	31.924 (0.002)	4.915 (0.568)	3013.45
Mix3, Token:	0.348 (0.677)			−0.439 (0.002)	−1.688 (0.01)	0.718 (0)	32.142 (0.211)	4.908 (0.609)	3012.54
Mix4, Token:	0.355 (0.078)			−0.201 (2.254)	−1.449 (0.312)	6.293 (71.046)	8.933 (0.001)	0 (*NaN*)	3562.17
Mix1, Value:	5.215 (6.114)			−0.448 (60.594)	−1.175 (1.464)	3.576 (479.438)	12.014 (0.001)		3526.02
Mix2, Value:	1.769 (0.145)			−0.388 (0.007)	−2.898 (0.003)	1.149 (0.021)	34.682 (0.043)	5.429 (0.31)	3053.88
Mix3, Value:	1.623 (0.247)			−0.385 (0.019)	−2.906 (0.005)	1.131 (0.057)	34.865 (0.001)	5.547 (2.315)	3053.32
Mix4, Value:	4.735 (0.605)			−0.118 (0.034)	−2.757 (0)	32.453 (9.144)	13.521 (0)	0.124 (*NaN*)	3521.2
Mix1, Contextual:	12.333 (1768.841)			−0.586 (459.226)	−1.062 (0.156)	4.377 (3455.141)	9.323 (49.096)		3569.01
Mix2, Contextual:	5.808 (139.885)			−0.231 (0.112)	−1.201 (8.927)	0.528 (0.242)	42.616 (812.487)	4.19 (135.05)	3076.64
Mix3, Contextual:	4.075 (0.19)			−0.555 (0)	−2.208 (0.003)	0.424 (0)	36.132 (0.025)	3.246 (0.24)	3074.77
Mix4, Contextual:	0.355 (0.078)			−0.201 (2.254)	−1.449 (0.312)	6.293 (71.046)	8.933 (0.001)	0 (*NaN*)	3562.17
Dictator games II (HJ06), Logit									
Mix1, Token:	0.143 (0.029)			0.589 (0)	−0.161 (1.051)	0.388 (0)	1.472 (0)		1694.4
Mix2, Token:	−1.528 (0.115)			0.272 (0)	−0.647 (0)	0.292 (0)	6.479 (0.001)	1.516 (0.056)	1623.38
Mix3, Token:	−1.528 (0.115)			0.272 (0)	−0.647 (0)	0.292 (0)	6.479 (0.001)	1.516 (0.056)	1623.38
Mix4, Token:	0.149 (0.038)			0.518 (0.009)	−0.432 (0.567)	0.378 (0.006)	1.974 (0.179)	1.683 (*NaN*)	1692.87
Mix1, Value:	1.791 (0.484)			0.459 (0.002)	−0.947 (0)	0.361 (0)	2.054 (0.75)		1690.96
Mix2, Value:	0.671 (0.596)			0.427 (0)	−0.698 (1.229)	0.372 (0)	6.265 (0)	1.555 (0.421)	1624.6

continued on next page

	λ	κ	τ	α	β	σ_α	σ_β	σ_λ	$\|LL\|$
Mix3, Value:	0.671 (0.596)			0.427 (0)	−0.698 (1.229)	0.372 (0)	6.265 (0)	1.555 (0.421)	1624.6
Mix4, Value:	1.548 (0.295)			0.419 (0)	−0.497 (0)	0.338 (0)	2.07 (0)	0.119 (*NaN*)	1689.67
Mix1, Contextual:	9.646 (21.216)			0.486 (0.002)	−0.486 (0)	0.401 (0.009)	2.183 (16.29)		1685.14
Mix2, Contextual:	2.56 (0.352)			0.428 (0)	−0.681 (0)	0.346 (0)	6.722 (0.002)	1.514 (0.226)	1621.27
Mix3, Contextual:	2.56 (0.348)			0.428 (0)	−0.681 (0)	0.346 (0)	6.722 (0.001)	1.514 (0.225)	1621.27
Mix4, Contextual:	0.149 (0.038)			0.518 (0.009)	−0.432 (0.567)	0.378 (0.006)	1.974 (0.179)	1.683 (*NaN*)	1692.87
Dictator games III (CHST07), Logit									
Mix1, Token:	0.025 (0.659)			0.257 (0.004)	−0.781 (147.783)	0.395 (0.007)	0.943 (282.071)		1034.58
Mix2, Token:	1.449 (1.106)			0.695 (*NaN*)	−0.617 (0.129)	0.002 (0)	11.541 (2.415)	99.997 (0)	910.14
Mix3, Token:	1.594 (0.832)			0.695 (*NaN*)	−0.617 (0)	0.002 (0)	11.541 (0)	100 (0)	910.1
Mix4, Token:	0.029 (0.007)			0.218 (0)	−0.485 (0)	0.378 (0)	0.883 (0)	1.353 (*NaN*)	1033.65
Mix1, Value:	0.174 (0.045)			0.161 (0.033)	−0.783 (0)	0.399 (0)	0.937 (0)		1046.47
Mix2, Value:	−0.365 (0.484)			0.522 (*NaN*)	−1.222 (0)	0.003 (0)	33.586 (0)	98.982 (25.392)	942.91
Mix3, Value:	−0.071 (0.284)			0.522 (*NaN*)	−1.222 (0)	0.003 (0)	33.58 (0.001)	98.852 (16.037)	942.98
Mix4, Value:	0.216 (0.072)			0.159 (0)	−0.33 (0)	0.474 (0)	0.724 (0)	0.122 (*NaN*)	1043.95
Mix1, Contextual:	9.538 (0.901)			0.162 (0)	−0.786 (0)	0.6 (0)	0.934 (0)		1036.67
Mix2, Contextual:	2.21 (1.546)			0.695 (*NaN*)	−0.617 (0)	0.002 (0)	11.541 (0)	88.386 (42.321)	909.49
Mix3, Contextual:	2.321 (0.793)			0.695 (*NaN*)	−0.617 (0)	0.002 (0)	11.541 (0)	88.386 (123.293)	909.46
Mix4, Contextual:	0.029 (0.007)			0.218 (0)	−0.485 (0)	0.378 (0)	0.883 (0)	1.353 (*NaN*)	1033.65
Dictator games with GUI (FKM07), Logit									
Mix1, Token:	0.041 (0.003)			0.317 (0)	−0.021 (0)	0.497 (0)	1.253 (0.007)		15682.07
Mix2, Token:	−1.985 (0.022)			0.064 (0)	0.243 (0)	0.839 (0)	1.662 (0)	2.019 (0.026)	15091.09
Mix3, Token:	−1.991 (0.034)			0.071 (0)	0.244 (0)	0.883 (0)	1.662 (0)	2.018 (0.016)	15090.96
Mix4, Token:	0.04 (0.002)			0.031 (0)	0.038 (0)	1.093 (0.001)	1.764 (0.003)	0.424 (*NaN*)	15684.71
Mix1, Value:	1.391 (0.028)			0.146 (0)	0.1 (0)	0.615 (0)	1.655 (0)		15625.66
Mix2, Value:	0.648 (0.019)			0.254 (0)	0.422 (0.001)	0.837 (0.001)	1.291 (0.004)	1.64 (0.079)	15100.02
Mix3, Value:	1.22 (0.319)			0.182 (0.002)	0.293 (0.002)	0.548 (0.004)	2.532 (0.002)	2.269 (0.228)	15095.86
Mix4, Value:	1.337 (0.137)			0.056 (0)	−0.026 (0)	0.653 (0.002)	1.688 (0)	0.068 (*NaN*)	15625.46
Mix1, Contextual:	0.042 (0.021)			0.317 (0)	−0.021 (0)	0.497 (0)	1.253 (0.015)		15682.32
Mix2, Contextual:	−1.766 (0.311)			0.26 (0.001)	0.38 (0)	0.494 (0)	2.345 (0)	1.711 (0.043)	15107.07
Mix3, Contextual:	−1.984 (0.027)			0.071 (0)	0.244 (0)	0.883 (0)	1.662 (0)	2.025 (0.039)	15090.94
Mix4, Contextual:	0.04 (0.002)			0.031 (0)	0.038 (0)	1.093 (0.001)	1.764 (0.003)	0.424 (*NaN*)	15684.71

Table 43: Estimated parameters of **FOCAL** models

	λ	κ	τ	α	β	σ_α	σ_β	σ_λ	$\|LL\|$
Dictator games I (AM02), FOCAL									

continued on next page

	λ	κ	τ	α	β	σ_α	σ_β	σ_λ	$\lvert LL \rvert$
Mix1, Token:	0.37 (0.039)	1.934 (0.064)		0.485 (0)	−0.112 (0.178)	0.323 (0)	2.122 (0)		2218.37
Mix2, Token:	−0.606 (0.094)	1.921 (0.065)		0.275 (0.002)	0.299 (0.002)	0.394 (0.001)	5.79 (0.018)	1.51 (0.087)	2133.21
Mix3, Token:	−0.762 (0.121)	0.62 (0.036)	0.129 (0.056)	0.161 (0.007)	0.172 (0.007)	0.37 (0.009)	7.345 (0.181)	1.446 (0.096)	2129.92
Mix4, Token:	0.37 (0.038)	0.684 (0.026)	0.18 (0.05)	0.331 (0)	0.084 (0.023)	0.27 (0)	2.495 (0.032)	1.884 (*NaN*)	2205.48
Mix1, Value:	4.044 (0.424)	1.912 (0.068)		0.417 (0.002)	−0.01 (0.383)	0.294 (0.001)	2.362 (0.007)		2209.94
Mix2, Value:	2.036 (0.168)	1.854 (0.07)		0.306 (0)	0.005 (0.045)	0.256 (0)	3.667 (0)	1.261 (0.228)	2151.72
Mix3, Value:	1.913 (0.131)	0.664 (0.033)	0.293 (0.076)	0.352 (0.055)	−0.022 (0.034)	0.324 (0.03)	3.745 (0.25)	1.095 (0.096)	2130.64
Mix4, Value:	4.079 (0.438)	0.683 (0.028)	0.221 (0.052)	0.391 (0)	−0.333 (0)	0.301 (0)	3.035 (0)	1.108 (*NaN*)	2189.48
Mix1, Contextual:	13.261 (3.316)	2.201 (0.11)		0.303 (0.025)	0.044 (0.004)	0.392 (0.027)	1.331 (0.092)		2261.48
Mix2, Contextual:	3.267 (0.135)	1.949 (0.066)		0.291 (0)	0.274 (0.006)	0.404 (0)	5.796 (0)	1.299 (0.147)	2160.64
Mix3, Contextual:	3.281 (0.127)	0.671 (0.032)	0.011 (0.142)	0.29 (0)	0.274 (0.007)	0.403 (0)	5.794 (0)	1.291 (0.15)	2160.14
Mix4, Contextual:	0.37 (0.038)	0.684 (0.026)	0.18 (0.05)	0.331 (0)	0.084 (0.023)	0.27 (0)	2.495 (0.032)	1.884 (*NaN*)	2205.48
Dictator games II (HJ06), FOCAL									
Mix1, Token:	0.165 (0.047)	1.522 (0.075)		0.498 (0.015)	−0.703 (2.853)	0.311 (0)	2.12 (3.512)		1234.39
Mix2, Token:	−1.757 (0.119)	1.541 (0.073)		0.387 (0)	−0.775 (0)	0.311 (0)	4.24 (0)	1.236 (0.122)	1191.74
Mix3, Token:	−1.436 (0.173)	0.42 (0.055)	0.209 (0.108)	0.413 (0)	−0.666 (0)	0.295 (0)	6.383 (0.001)	1.523 (0.141)	1187.59
Mix4, Token:	0.173 (0.03)	0.443 (0.053)	0.322 (0.092)	0.378 (0)	−0.185 (0.117)	0.273 (0)	1.56 (0)	0.987 (*NaN*)	1225.06
Mix1, Value:	1.662 (0.315)	1.513 (0.074)		0.32 (0)	−0.628 (0.406)	0.255 (0)	2.009 (0)		1230.2
Mix2, Value:	0.865 (0.096)	1.546 (0.075)		0.326 (0)	−0.628 (0)	0.275 (0)	2.552 (0)	1.292 (0.052)	1187.04
Mix3, Value:	0.702 (0.195)	0.422 (0.057)	0.192 (0.094)	0.429 (0)	−0.721 (0)	0.361 (0)	6.488 (0.009)	1.569 (0.119)	1183.04
Mix4, Value:	1.748 (0.301)	0.42 (0.065)	0.288 (0.109)	0.393 (0)	−0.366 (0.405)	0.256 (0)	1.742 (0)	1.459 (*NaN*)	1221.24
Mix1, Contextual:	10.512 (1.683)	1.535 (0.076)		0.306 (0)	−0.551 (0)	0.282 (0)	1.936 (0)		1236.12
Mix2, Contextual:	2.797 (0.151)	1.561 (0.076)		0.33 (0)	−0.602 (0)	0.281 (0)	2.57 (0)	1.2 (0.131)	1188.47
Mix3, Contextual:	2.757 (0.107)	0.412 (0.064)	0.218 (0.102)	0.395 (0)	−0.669 (0)	0.327 (0)	6.664 (0.001)	1.309 (0.141)	1180.58
Mix4, Contextual:	0.173 (0.03)	0.443 (0.053)	0.322 (0.092)	0.378 (0)	−0.185 (0.117)	0.273 (0)	1.56 (0)	0.987 (*NaN*)	1225.06
Dictator games III (CHST07), FOCAL									
Mix1, Token:	0.031 (0.006)	2.173 (0.087)		1 (0.015)	−0.15 (0.228)	0.69 (0)	0 (0.015)		422.11
Mix2, Token:	−2.634 (0.272)	2.273 (0.108)		1 (13.071)	−0.141 (0.002)	0.662 (0.017)	0.008 (0.001)	1.263 (0.123)	409.51
Mix3, Token:	−2.661 (0.236)	0.821 (0.048)	0 (37.496)	1 (37.523)	−0.144 (0)	0.667 (0)	0.007 (0)	1.276 (0.126)	409.45
Mix4, Token:	0.031 (0.005)	0.777 (0.04)	0.008 (0.004)	1 (51.999)	−0.149 (0)	0.686 (0)	0.001 (0)	0.118 (*NaN*)	422.14
Mix1, Value:	0.193 (0.036)	2.176 (0.086)		0.986 (0.629)	−0.198 (0)	0.677 (0.255)	0.001 (0.004)		420.24
Mix2, Value:	−0.93 (0.283)	2.272 (0.11)		0.933 (0.022)	−0.198 (0)	0.651 (0)	0 (0.052)	1.121 (0.141)	408.37
Mix3, Value:	−0.913 (0.271)	0.817 (0.048)	0.001 (0.007)	0.944 (0)	−0.198 (0)	0.656 (0)	0.002 (0)	1.176 (0.138)	408.22
Mix4, Value:	0.192 (0.037)	0.779 (0.04)	0.003 (0.013)	0.734 (0.266)	−0.198 (0)	0.59 (0)	0.001 (0.301)	2.325 (*NaN*)	420.61
Mix1, Contextual:	14.864 (1.557)	2.189 (0.083)		0.994 (0.158)	−0.15 (0.006)	0.711 (0.038)	0 (0.002)		407.99
Mix2, Contextual:	3.407 (0.364)	2.218 (0.1)		1 (0.017)	−0.155 (0)	0.681 (0.01)	0.005 (0)	0.884 (0.285)	406.54
Mix3, Contextual:	3.372 (0.343)	0.797 (0.045)	0 (0.006)	0.887 (0.203)	−0.148 (0)	0.653 (0.08)	0.005 (0.001)	0.867 (0.285)	406.55

continued on next page

	λ	κ	τ	α	β	σ_α	σ_β	σ_λ	$\|LL\|$
Mix4, Contextual:	0.031 (0.005)	0.777 (0.04)	0.008 (0.004)	1 (51.999)	−0.149 (0)	0.686 (0)	0.001 (0)	0.118 (*NaN*)	422.14
Dictator games with GUI (FKM07), FOCAL									
Mix1, Token:	0.041 (0.001)	0.105 (0.005)		−0.002 (0)	0.185 (0.002)	0.437 (0.01)	1.533 (0.01)		15668.79
Mix2, Token:	−1.939 (0.013)	0.113 (0.001)		0.065 (0)	0.243 (0)	0.551 (0)	1.584 (0.001)	1.893 (0.02)	15082.56
Mix3, Token:	−1.925 (0.032)	−2.973 (0.08)	1.334 (0.018)	0.082 (0)	0.257 (0)	0.625 (0)	1.599 (0.002)	1.88 (0.027)	15070.9
Mix4, Token:	0.041 (0.002)	−3.034 (0.236)	1.622 (0.051)	0.112 (0)	−0.065 (0)	0.425 (0)	1.286 (0.002)	1.409 (*NaN*)	15600.73
Mix1, Value:	1.368 (0.315)	0.138 (0.017)		0.063 (0.002)	−0.021 (0)	0.51 (0.005)	1.978 (0.018)		15592.67
Mix2, Value:	0.626 (0.018)	0.104 (0.002)		0.236 (0.002)	0.424 (0.001)	0.806 (0.002)	1.29 (0.002)	1.634 (0.029)	15078.48
Mix3, Value:	0.633 (0.011)	−2.223 (0.009)	0 (0)	0.26 (0)	0.425 (0)	0.847 (0.001)	1.289 (0)	1.637 (0.017)	15077.99
Mix4, Value:	1.362 (0.036)	−2.536 (0.27)	2.642 (0.285)	0.117 (0.001)	−0.189 (0)	0.386 (0)	2.141 (0)	0.683 (*NaN*)	15506.86
Mix1, Contextual:	11.796 (0.144)	0 (0.001)		0.204 (0)	−0.018 (0.001)	0.51 (0.001)	1.903 (0.026)		15623.08
Mix2, Contextual:	3.933 (0.027)	0.013 (0)		0.114 (0)	0.233 (0)	0.273 (0)	1.392 (0.001)	1.385 (0.018)	15132
Mix3, Contextual:	3.591 (0.043)	−3.564 (0.082)	1.03 (0.011)	0.319 (0.001)	0.175 (0)	0.386 (0.001)	1.025 (0)	1.596 (0.019)	15088.74
Mix4, Contextual:	0.041 (0.002)	−3.034 (0.236)	1.622 (0.051)	0.112 (0)	−0.065 (0)	0.425 (0)	1.286 (0.002)	1.409 (*NaN*)	15600.73

Table 44: Estimated parameters of **PALM**

	λ	κ	τ	α	β	σ_α	σ_β	σ_λ	$\|LL\|$
Dictator games I (AM02), PALM									
Mix1, Token:	0.262 (0.057)	1 (*NaN*)		−0.68 (*NaN*)	0.018 (*NaN*)	6.117 (*NaN*)	5.408 (172.782)		3419.12
Mix2, Token:	−0.929 (0.173)	1 (10.525)		−0.459 (0.002)	0.254 (0.027)	0.274 (0.001)	77.277 (0.007)	4.534 (0.424)	2879.03
Mix3, Token:	−0.935 (0.148)	0.881 (*NaN*)	0.086 (*NaN*)	−0.457 (0)	0.252 (0.087)	0.274 (0)	77.507 (0.017)	4.537 (1.418)	2877.26
Mix4, Token:	0.403 (0.084)	2.319 (*NaN*)	0.496 (*NaN*)	−0.788 (9.074)	−0.376 (0.214)	3.199 (38.657)	11.699 (0)	6.006 (*NaN*)	3361.19
Mix1, Value:	5.025 (24.806)	1 (*NaN*)		−0.705 (*NaN*)	0.129 (*NaN*)	2.909 (*NaN*)	11.921 (88.384)		3328.84
Mix2, Value:	1.871 (0.306)	1 (7083385084727.88)		−0.439 (*NaN*)	0.236 (0.004)	0.595 (*NaN*)	76.655 (0.009)	4.76 (0.733)	2919.49
Mix3, Value:	2.078 (1.118)	1.764 (*NaN*)	0.123 (*NaN*)	−0.308 (1.734)	0.15 (0)	1.948 (5.716)	99.925 (0.03)	3.911 (2.827)	2894.94
Mix4, Value:	4.616 (0.73)	2.456 (*NaN*)	0.542 (*NaN*)	−0.688 (0.017)	−0.331 (0)	2.03 (0.051)	14.11 (0)	5.454 (*NaN*)	3332.49
Mix1, Contextual:	14.645 (5.924)	1 (*NaN*)		−0.628 (*NaN*)	0.13 (*NaN*)	5.352 (*NaN*)	11.255 (1.344)		3407.15
Mix2, Contextual:	4.625 (0.277)	1 (0.001)		−0.336 (0)	0.058 (0)	0.292 (0)	86.847 (0.135)	3.324 (0.252)	2925.49
Mix3, Contextual:	4.804 (0.209)	2.28 (*NaN*)	0.084 (*NaN*)	−0.265 (0)	0.12 (0.011)	0.233 (0)	68.753 (0.002)	2.661 (0.162)	2908.1
Mix4, Contextual:	0.403 (0.084)	2.319 (*NaN*)	0.496 (*NaN*)	−0.788 (9.074)	−0.376 (0.214)	3.199 (38.657)	11.699 (0)	6.006 (*NaN*)	3361.19
Dictator games II (HJ06), PALM									
Mix1, Token:	0.152 (0.032)	1 (11067317877451.7)		0.623 (*NaN*)	−0.165 (0.614)	0.395 (*NaN*)	1.5 (*NaN*)		1631.98
Mix2, Token:	−1.726 (0.168)	1 (*NaN*)		0.252 (*NaN*)	−0.935 (*NaN*)	0.252 (*NaN*)	6.037 (0.014)	2.347 (0.195)	1576.44
Mix3, Token:	−1.537 (0.056)	0.432 (*NaN*)	0.105 (*NaN*)	0.272 (0.001)	−0.649 (0.003)	0.292 (0)	6.472 (0.014)	1.514 (0.055)	1560.99
Mix4, Token:	0.149 (0.033)	0.915 (*NaN*)	0.064 (*NaN*)	0.518 (0.015)	−0.432 (0.312)	0.378 (0.011)	1.974 (0.137)	1.683 (*NaN*)	1630.24
Mix1, Value:	1.548 (0.614)	1 (*NaN*)		0.363 (*NaN*)	−0.389 (2.384)	0.294 (*NaN*)	1.524 (*NaN*)		1628.11

continued on next page

	λ	κ	τ	α	β	σ_α	σ_β	σ_λ	$\|LL\|$
Mix2, Value:	0.896 (0.198)	1 (*NaN*)		0.253 (*NaN*)	−0.348 (0.339)	0.291 (*NaN*)	3.068 (*NaN*)	1.506 (0.589)	1574.14
Mix3, Value:	0.671 (0.484)	0.426 (*NaN*)	0.097 (*NaN*)	0.427 (0)	−0.698 (0.825)	0.372 (0)	6.265 (0)	1.555 (0.148)	1562.31
Mix4, Value:	1.548 (0.502)	1.061 (*NaN*)	0.077 (*NaN*)	0.419 (0)	−0.497 (1.936)	0.337 (0)	2.069 (0)	0.119 (*NaN*)	1626.23
Mix1, Contextual:	9.758 (12.221)	1 (*NaN*)		0.589 (*NaN*)	−0.157 (*NaN*)	0.389 (*NaN*)	1.446 (1.311)		1624.3
Mix2, Contextual:	2.744 (0.142)	1 (*NaN*)		0.262 (*NaN*)	−0.419 (0.044)	0.244 (*NaN*)	4.137 (*NaN*)	1.295 (0.115)	1564.54
Mix3, Contextual:	2.552 (0.278)	0.404 (*NaN*)	0.114 (*NaN*)	0.427 (0)	−0.684 (0)	0.344 (0)	6.716 (0.081)	1.52 (0.232)	1559.73
Mix4, Contextual:	0.149 (0.033)	0.915 (*NaN*)	0.064 (*NaN*)	0.518 (0.015)	−0.432 (0.312)	0.378 (0.011)	1.974 (0.137)	1.683 (*NaN*)	1630.24
Dictator games III (CHST07), PALM									
Mix1, Token:	0.025 (0.172)	0.853 (15.502)		0.257 (0)	−0.781 (7.791)	0.395 (0.001)	0.943 (6.573)		1011
Mix2, Token:	1.376 (1.709)	1 (1.877)		0.695 (*NaN*)	−0.617 (0)	0.002 (0)	11.541 (0)	99.995 (0)	889.55
Mix3, Token:	1.594 (1.021)	0.11 (*NaN*)	0.01 (*NaN*)	0.695 (*NaN*)	−0.617 (0)	0.002 (0)	11.541 (0)	99.999 (0)	889.54
Mix4, Token:	0.029 (0.006)	0.382 (*NaN*)	0.107 (*NaN*)	0.218 (0)	−0.485 (0)	0.378 (0)	0.883 (0)	1.353 (*NaN*)	1005.45
Mix1, Value:	0.174 (0.058)	0.853 (1.081)		0.161 (0.034)	−0.783 (0.122)	0.399 (0)	0.937 (0)		1021.82
Mix2, Value:	−0.171 (0.312)	1 (*NaN*)		0.522 (*NaN*)	−1.222 (*NaN*)	0.003 (*NaN*)	33.584 (*NaN*)	98.732 (22.138)	921.06
Mix3, Value:	−0.061 (0.356)	0.09 (*NaN*)	0.01 (*NaN*)	0.522 (*NaN*)	−1.222 (0)	0.003 (0)	33.579 (0.001)	99.028 (21.316)	921.04
Mix4, Value:	0.217 (0.076)	0.94 (*NaN*)	0.084 (*NaN*)	0.159 (0)	−0.329 (0)	0.472 (0)	0.725 (0)	0.123 (*NaN*)	1014.83
Mix1, Contextual:	9.508 (0.799)	0.893 (0.555)		0.162 (0)	−0.786 (0)	0.599 (0)	0.933 (0)		1005.91
Mix2, Contextual:	2.321 (0.838)	1 (2.156)		0.695 (*NaN*)	−0.617 (0)	0.002 (0)	11.541 (0)	88.385 (38.981)	888.93
Mix3, Contextual:	2.321 (0.844)	0.11 (*NaN*)	0.01 (*NaN*)	0.695 (*NaN*)	−0.617 (0)	0.002 (0)	11.541 (0)	88.386 (39.988)	888.92
Mix4, Contextual:	0.029 (0.006)	0.382 (*NaN*)	0.107 (*NaN*)	0.218 (0)	−0.485 (0)	0.378 (0)	0.883 (0)	1.353 (*NaN*)	1005.45
Dictator games with GUI (FKM07), PALM									
Mix1, Token:	0.041 (0.005)	0.14 (0.003)		0.318 (0)	−0.021 (0)	0.497 (0)	1.258 (0.017)		15680.58
Mix2, Token:	−2 (0.016)	0.339 (0.008)		0.064 (0)	0.243 (0)	0.836 (0)	1.663 (0)	2.025 (0.025)	15087.13
Mix3, Token:	−1.996 (0.06)	−0.535 (0.005)	0.009 (0)	0.063 (0)	0.243 (0)	0.833 (0)	1.663 (0)	2.052 (0.032)	15086.44
Mix4, Token:	0.041 (0.003)	−1.932 (0.027)	4.28 (0.153)	0.138 (0)	−0.046 (0)	0.657 (0)	1.295 (0)	1.278 (*NaN*)	15676.36
Mix1, Value:	1.36 (0.145)	0.704 (0.013)		0.264 (0.001)	0.099 (0)	0.648 (0)	1.854 (0)		15613.92
Mix2, Value:	0.629 (0.017)	0.789 (0.011)		0.266 (0)	0.428 (0.001)	0.856 (0.001)	1.289 (0.001)	1.625 (0.032)	15087.88
Mix3, Value:	1.18 (0.033)	−0.06 (0.004)	0.209 (0.007)	0.2 (0)	0.302 (0)	0.576 (0)	2.546 (0.001)	2.258 (0.111)	15082.37
Mix4, Value:	1.437 (0.626)	−0.386 (0.058)	4.445 (0.519)	0.206 (0)	−0.171 (0)	0.393 (0)	2.199 (0.005)	0.747 (*NaN*)	15605.96
Mix1, Contextual:	11.796 (0.061)	0 (0)		0.204 (0)	−0.018 (0)	0.51 (0)	1.903 (0.01)		15623.06
Mix2, Contextual:	3.934 (0.045)	0.013 (0.001)		0.114 (0)	0.233 (0)	0.273 (0)	1.393 (0)	1.385 (0.032)	15133.63
Mix3, Contextual:	3.562 (0.06)	−3.711 (0.444)	0.49 (0.091)	0.316 (0.001)	0.175 (0.001)	0.384 (0.001)	1.01 (0.002)	1.555 (0.073)	15095.95
Mix4, Contextual:	0.041 (0.003)	−1.932 (0.027)	4.28 (0.153)	0.138 (0)	−0.046 (0)	0.657 (0)	1.295 (0)	1.278 (*NaN*)	15676.36

Table 45: Estimated parameters of **OGEV** models

	λ	κ	τ	α	β	σ_α	σ_β	σ_λ	$\|LL\|$
Dictator games I (AM02), OGEV									
Mix1, Token:	0.075 (0.418)	0.007 (0.405)		0.077 (0)	−0.516 (0)	2.433 (1405.889)	9.993 (7.646)		3125.69
Mix2, Token:	−2.18 (0.217)	0.001 (0.001)		−0.081 (0.005)	−0.099 (0.006)	3.995 (0.386)	86.455 (0)	3.052 (0.536)	2778.79
Mix3, Token:	−1.452 (0.119)	−3.714 (0.202)	0.039 (0.17)	−0.56 (0.021)	−0.164 (0.009)	0.561 (0.021)	85.802 (0)	3.199 (0.172)	2759.77
Mix4, Token:	0.076 (0.013)	−4.93 (0.931)	2.419 (1.156)	−0.523 (0)	−0.373 (0.002)	0.924 (0.001)	45.421 (0.098)	2.956 (*NaN*)	3104.16
Mix1, Value:	0.356 (0.111)	0.003 (0.002)		0.11 (0.538)	−0.389 (0.173)	1.527 (2.435)	86.018 (38.173)		2944.31
Mix2, Value:	0.117 (0.182)	0.001 (0.001)		−0.065 (0)	−0.116 (0)	4.702 (0)	84.829 (0.002)	2.887 (0.248)	2776.2
Mix3, Value:	0.823 (0.098)	−3.705 (0.271)	0.13 (0.19)	−0.473 (0)	−0.227 (0.019)	0.498 (0)	85.63 (0.076)	2.833 (0.158)	2770.38
Mix4, Value:	1.408 (0.363)	−4.509 (0.34)	0.545 (0.072)	−0.309 (0.098)	−0.239 (0.015)	8.019 (2.777)	34.195 (0.001)	1.812 (*NaN*)	3108.47
Mix1, Contextual:	2.989 (0.359)	0.002 (0)		−0.019 (0.004)	−0.141 (0)	1.629 (0.302)	67.917 (0.004)		2977.66
Mix2, Contextual:	3.356 (13.034)	0.001 (0.114)		−0.078 (0.169)	−0.093 (0)	5.962 (19.831)	68.951 (0.091)	3.934 (45.256)	2796.25
Mix3, Contextual:	4.039 (4.146)	−3.795 (1.687)	0.372 (3.074)	−0.69 (0.168)	−0.12 (1.109)	0.494 (0.108)	86.792 (105.46)	3.36 (4.629)	2766.57
Mix4, Contextual:	0.076 (0.013)	−4.93 (0.931)	2.419 (1.156)	−0.523 (0)	−0.373 (0.002)	0.924 (0.001)	45.421 (0.098)	2.956 (*NaN*)	3104.16
Dictator games II (HJ06), OGEV									
Mix1, Token:	0.108 (0.02)	0.021 (0.102)		0.392 (0.004)	−0.876 (0.008)	0.27 (0)	2.722 (0)		1664.94
Mix2, Token:	−2.026 (0.256)	0.001 (0.023)		0.274 (0)	0.012 (0)	0.492 (0)	9.785 (0)	1.693 (0.143)	1620.65
Mix3, Token:	−2.023 (0.234)	−0.054 (2.463)	3.725 (2.172)	0.234 (0)	−0.596 (0.015)	0.327 (0)	6.182 (0)	1.578 (0.134)	1606.79
Mix4, Token:	0.124 (0.023)	−2.125 (0.32)	5.495 (0.565)	0.377 (0)	−0.267 (0.001)	0.264 (0)	1.837 (0)	0.078 (*NaN*)	1633.41
Mix1, Value:	0.895 (0.168)	0.032 (0.075)		0.474 (0.326)	−0.85 (0.583)	0.311 (0.001)	3.001 (0)		1657.82
Mix2, Value:	−0.167 (0.217)	0.001 (0.161)		0.2 (0)	0.028 (0.005)	0.493 (0.001)	7.935 (0.048)	1.731 (0.171)	1614.38
Mix3, Value:	0.326 (0.165)	−1.81 (0.507)	1.979 (0.503)	0.473 (0)	−0.715 (0)	0.425 (0)	6.341 (0.001)	1.604 (0.095)	1603.7
Mix4, Value:	1.181 (0.201)	−2.882 (0.358)	3.661 (0.409)	0.386 (0)	−0.356 (0.001)	0.293 (0)	1.867 (0)	0.328 (*NaN*)	1632.52
Mix1, Contextual:	4.441 (0.625)	0.069 (0.027)		0.239 (0)	−1.039 (0.156)	0.301 (0)	2.744 (0.183)		1659.18
Mix2, Contextual:	2.554 (0.267)	0.406 (0.655)		0.428 (0)	−0.677 (0.001)	0.345 (0)	6.664 (0.007)	1.492 (0.21)	1618.54
Mix3, Contextual:	2.448 (0.116)	−0.255 (0.672)	3.811 (0.981)	0.233 (0.001)	−0.661 (0.003)	0.486 (0.001)	6.737 (0.013)	1.361 (0.086)	1606.71
Mix4, Contextual:	0.124 (0.023)	−2.125 (0.32)	5.495 (0.565)	0.377 (0)	−0.267 (0.001)	0.264 (0)	1.837 (0)	0.078 (*NaN*)	1633.41
Dictator games III (CHST07), OGEV									
Mix1, Token:	0.062 (0.015)	0 (2.959)		0.414 (0.001)	−0.396 (0.004)	0.173 (0)	1.385 (0.004)		975.66
Mix2, Token:	4.613 (2.077)	0.001 (0)		0.73 (*NaN*)	−0.617 (0)	0.002 (*NaN*)	11.541 (0)	80.292 (19.355)	881.27
Mix3, Token:	2.197 (2.843)	−4.605 (1.937)	0.11 (0.904)	0.73 (*NaN*)	−0.617 (0)	0.002 (0)	11.541 (0)	80.152 (33.612)	898.97
Mix4, Token:	0.038 (0.018)	−10.647 (0.784)	0.009 (0.112)	0.211 (0)	−0.399 (0)	0.435 (0)	1.512 (0)	1.199 (*NaN*)	1024.67
Mix1, Value:	0.193 (0.04)	0.001 (0)		0.384 (0)	−0.672 (0)	0.295 (0)	1.651 (0)		989.7
Mix2, Value:	−0.071 (0.49)	0.001 (0)		0.634 (*NaN*)	−1.223 (0)	0.003 (0.002)	33.594 (0)	44.223 (14.498)	911.53
Mix3, Value:	−0.171 (34.316)	−4.254 (6.04)	0.02 (6.273)	0.604 (*NaN*)	−1.222 (88.177)	0.003 (0)	33.592 (2422.988)	88.983 (1522.519)	932.27
Mix4, Value:	0.216 (0.072)	0.94 (*NaN*)	0.084 (*NaN*)	0.159 (0)	−0.33 (0)	0.474 (0)	0.724 (0)	0.122 (*NaN*)	1043.95

continued on next page

	λ	κ	τ	α	β	σ_α	σ_β	σ_λ	$\lvert LL \rvert$
Mix1, Contextual:	10.66 (1.64)	0.001 (0)		0.264 (0)	−0.447 (0.036)	0.197 (0)	1.583 (0)		990.73
Mix2, Contextual:	2.431 (1.231)	0.001 (0)		0.73 (*NaN*)	−0.617 (0)	0.002 (*NaN*)	11.541 (0)	71.028 (32.966)	876.75
Mix3, Contextual:	3.315 (0.569)	−4.605 (1.482)	0.11 (1.636)	0.729 (*NaN*)	−0.617 (0)	0.002 (0)	11.541 (0)	85.257 (63.87)	898.14
Mix4, Contextual:	0.038 (0.018)	−10.647 (0.784)	0.009 (0.112)	0.211 (0)	−0.399 (0)	0.435 (0)	1.512 (0)	1.199 (*NaN*)	1024.67
Dictator games with GUI (FKM07), OGEV									
Mix1, Token:	0.043 (0.032)	0.356 (0.014)		0.243 (0)	−0.115 (0)	0.344 (0)	1.83 (0.001)		15644.38
Mix2, Token:	−2.005 (0.019)	0.459 (0.024)		0.067 (0)	0.25 (0)	0.56 (0)	1.573 (0)	1.909 (0.041)	15069.34
Mix3, Token:	−3.009 (0.05)	−0.003 (0)	2.217 (0.009)	0.245 (0)	−0.171 (0)	0.3 (0)	1.672 (0)	1.061 (0.014)	15054.77
Mix4, Token:	0.04 (0.002)	−1.935 (0.065)	2.705 (0.07)	0.035 (0)	0.029 (0)	1.102 (0)	1.748 (0)	0.424 (*NaN*)	15523.91
Mix1, Value:	1.352 (0.27)	0.31 (0.019)		0.16 (0)	0.256 (0.001)	0.395 (0)	2.216 (0.002)		15544.24
Mix2, Value:	0.598 (0.029)	0.57 (0.009)		0.167 (0)	0.421 (0)	0.67 (0)	1.281 (0)	1.592 (0.101)	15073.17
Mix3, Value:	1.125 (0.023)	−0.954 (0.012)	1.337 (0.016)	0.275 (0)	0.274 (0)	0.707 (0)	2.442 (0)	2.474 (0.06)	15014.72
Mix4, Value:	1.301 (0.004)	−2.415 (0.024)	2.808 (0.01)	0.127 (0)	−0.173 (0)	0.366 (0)	1.927 (0)	0.711 (*NaN*)	15430.79
Mix1, Contextual:	11.383 (0.036)	0.999 (0.003)		0.107 (0)	−0.187 (0)	0.471 (0)	1.409 (0)		15622.59
Mix2, Contextual:	3.898 (0.046)	1 (322982.404)		0.094 (0)	0.239 (0)	0.294 (0)	1.41 (0.001)	1.397 (0.043)	15127.77
Mix3, Contextual:	3.694 (0.22)	−2.743 (0.08)	3.742 (0.095)	0.171 (0)	0.189 (0)	0.339 (0)	1.023 (0)	1.331 (0.01)	15047.94
Mix4, Contextual:	0.04 (0.002)	−1.935 (0.065)	2.705 (0.07)	0.035 (0)	0.029 (0)	1.102 (0)	1.748 (0)	0.424 (*NaN*)	15523.91