RATIONALITY & COMPETITION
CRC TRR 190

Discrete Choice with Presentation Effects

Yves Breitmoser (Humboldt University Berlin)

Discussion Paper No. 35

June 6, 2017

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

Discrete choice with presentation effects

Yves Breitmoser*
Humboldt University Berlin

June 5, 2017

Abstract

Experimenters have to make theoretically irrelevant decisions concerning user interfaces and ordering or labeling of options. Such presentation decisions affect behavior and cause results to appear contradictory across experiments, obstructing utility estimation and policy recommendations. The present paper derives a model of choice allowing analysts to control for both presentation effects and stochastic errors in econometric analyses. I test the model in a comprehensive re-analysis of dictator game experiments. Controlling for presentation effects, preference estimates are consistent across experiments and predictive out-of-sample, highlighting the fundamental role of presentation for choice, and this notwithstanding the possibility of reliable estimation and prediction.

JEL–Code: C10, C90
Keywords: discrete choice, presentation effects, utility estimation, counterfactual predictions, laboratory experiment

*I thank Friedel Bolle, Annemarie Gronau, Matthias Lang, Nick Netzer, Martin Pollrich, Sebastian Schweighofer-Kodritsch, Justin Valasek, Roel van Veldhuizen, Georg Weizsäcker and audiences in Berlin (BERA workshop), Kreuzlingen (THEEM 2016), Heidelberg, and Schwanenwerder (CRC retreat) for many helpful comments. Support of the DFG (BR 4648/1 and CRC TRR 190) is greatly appreciated. Address: Spandauer Str. 1, 10099 Berlin, Germany, email: yves.breitmoser@hu-berlin.de, phone: +49 30 2093 99408.

1 Introduction

Economic studies analyze individual choice in order to understand preferences, amongst others social, time and risk preferences. These analyses make for a large and important literature, as understanding preferences is required as basis for theoretical analyses and policy recommendations. Estimated preferences appear to be contradictory across studies, however, which drastically limits their usefulness for making counterfactual predictions and policy recommendations. This inconsistency is particularly prevalent in studies of social preferences. As I show here, a major reason for this inconsistency appears to be that presentation of options affects choice. Presentation matters due to default effects (McKenzie et al., 2006; Dinner et al., 2011), left-digit effects (Poltrock and Schwartz, 1984; Lacetera et al., 2012), round-number effects (Heitjan and Rubin, 1991; Manski and Molinari, 2010), and positioning effects (Dean, 1980; Miller and Krosnick, 1998; Feenberg et al., 2017). Absent a model of such presentation effects that would allow researchers to control for them, they induce biased and inconsistent utility estimates. This implies the perceived incompatibility of observations and may also be a major reason for the failure to reach a consensus on (social) preference theories.

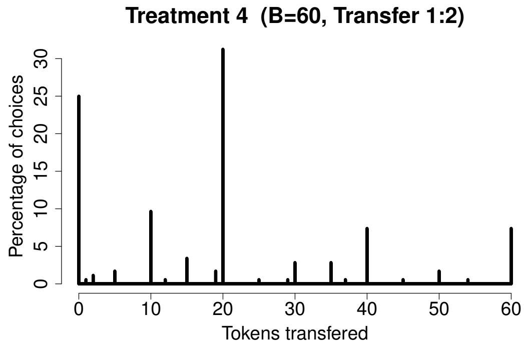
To fix ideas, consider an experimenter designing a dictator game experiment to test a model of social preferences. In a dictator game, player 1 ("dictator") chooses how many tokens, $x \in \{0,\dots,B\}$, to transfer to player 2 ("recipient"). Whatever model the experimenter seeks to test, some theoretically irrelevant decisions have to be made to run the experiment. I refer to them as presentation decisions. For example, the value of the cake to be redistributed may be fixed and the experimenter has to set the total number of "slices" B to run the experiment. By all received theories, the choice of B is theoretically irrelevant in that the budget share transferred by player 1 is independent of B (aside from discreteness). Yet, if $B = 20$, then the equal split results from a transfer of 10 tokens, and if say $B = 25$, then the equal split is not attainable at all, let alone by choosing a round number.

Such presentation decisions have striking behavioral effects, as Figure 1 illustrates. The figure revisits results from dictator game experiments that are behaviorally equivalent in the sense that if behavior depends solely on preferences, and preferences are functions of payoff profiles (as assumed in models of social and risk preferences), then the observed choice patterns should be statistically indistinguishable. Yet, choice patterns depend substantially on the roundedness of the equal-split option (Figures 1a and 1b), patterns are entirely different if choices are entered in graphical user interfaces (Figure 1c) and qualitative results such as comparative statics are contradictory (see Figure 1d)—all of which results in utility estimates that are inconsistent across studies (shown below). The implications are substantial. Since comparative statics depend on presentation, experimental studies cannot measure comparative statics (without controlling for presentation) and experimental results cannot be taken at face value. Since utilities are inconsistently estimated, social preference models are not predictive, evaluating preference models by comparing results from different experiments is futile, and convergence of social preference theory is put out of reach. Such concerns regularly surface in critiques of experimental and behavioral economics, and to address them, we need to control for presentation effects.

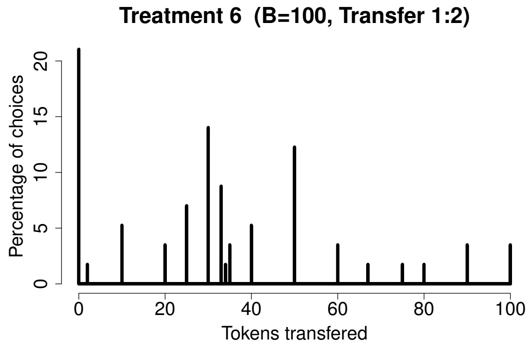
The present paper derives a model of presentation effects, applies it to standard data sets (including those in Figure 1), and shows that it allows to effectively factor out and con-

Figure 1: Choice in dictator games across experiments

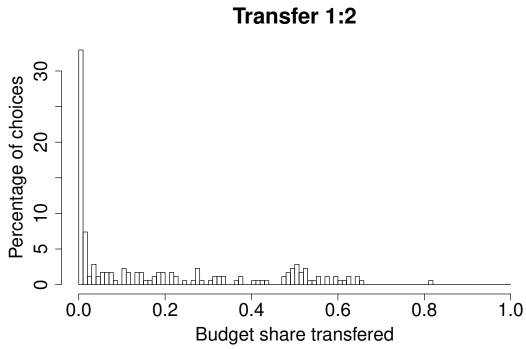
(a) AM02 Treatment 4



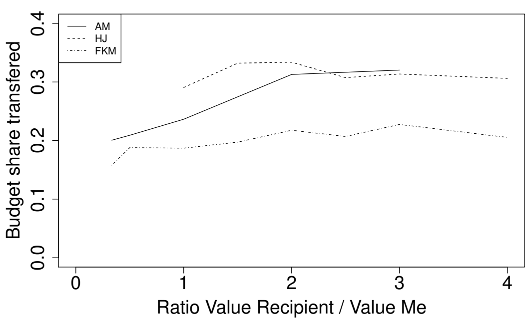
(b) HJ06 Treatment 6



(c) FKM07 “Treatment” 7



(d) Comparative statics



Note: All data sets are introduced in detail in Section 3. The plots here show behavior in games where each token is worth one point to the dictator and two points to the recipient. In Andreoni and Miller (2002, AM02), the equal split is a round number (transferring 20 tokens), in Harrison and Johnson (2006, HJ06) it is not a round number (transferring 33–34 tokens), and Fisman et al. (2007, FKM07) use a graphical user interface that does not directly show the number of tokens redistributed. Further plots are provided as supplement. Figure 1d shows the comparative statics of the budget share transferred by the dictator in the transfer rate, ranging from $1/4$ (highly taxed transfers) to $4/1$ (highly subsidized transfers).

trol for presentation effects in analyses. These results are fairly positive in nature. While behavior and social, time, and risk preferences appear to differ a lot across experiments, this paper provides both an analytical framework and econometric evidence suggesting that neither the preference theories nor experimental measurement as such are necessarily inadequate. Instead, current measures are being confounded by presentation effects, which we need to filter out, but doing so, we can get much further with our previous theories and measures than we previously thought.

Section 2 provides a short and less technical overview of approach and results. Section 3 contains the theoretical derivation and discussion of the focal choice adjusted logit (Focal) model. Section 4 describes data sets and approach to the empirical analysis, testing both adequacy and usefulness of the Focal model. Section 5 presents the econometric results and Section 6 concludes. The appendix contains relegated proofs and definitions, the online supplement contains reports on a number of robustness checks.

2 Overview of approach and results

In order to formally capture presentation effects, let us assume that decision maker DM seeks to maximize some utility function, but DM gets distracted by the presentation of the choice task and is "tempted" to choose options that are particularly convenient or focal. Gul and Pesendorfer (2001) introduce a general model of set-based preferences, allowing to capture the notion that DM's evaluation of a set may be lower than her evaluation of the best element in the set—choice as such is a non-trivial task and extraneous influences can obstruct finding and choosing the best option. These influences may stem from presentation of options, from temptation of options, from computational or practical difficulties, and so on. Gul and Pesendorfer show that DM's decision can be captured as an attempt to maximize a sum $u_x + v_x$ over the available options $x \in B$. The existence of corresponding indices u and v follows from Gul and Pesendorfer's axioms, u_x captures the utility associated with option x and v_x captures the "obstruction" emanating from option x. Considering the applications listed in this paper, I propose to interpret v_x as the relative focality of option x, which is a function of the presentation of options, and denote it as ϕ_x.

To apply this insight to experimental analyses, we need to extend Gul and Pesendorfer's results to stochastic choice. The main challenge is that, given the range of conceivable assumptions about locus and distribution of noise in the decision making process, there are myriad ways of doing so, and the econometric results eventually will depend on which assumptions we make (Hey, 2005; Wilcox, 2008). I address this challenge by deriving a model of stochastic choice without restricting locus and distribution of noise. Taking Gul and Pesendorfer's main result, existence of distinct factors u and ϕ, as foundation, I translate their result that DM maximizes $u_x + \phi_x$ into four invariance statements about choice and merely add the "positivity" axiom (all options have positive probability) to inject stochastic choice. Note that positivity does not put a restriction on locus or distribution of noise. The invariance statements are that choice satisfies independence of irrelevant alternatives (IIA), that choice depends solely on u and ϕ, and that it is invariant to translation of u and ϕ. These "axioms" represent $u_x + \phi_x$ in the sense that they yield $u_x + \phi_x$ under rational choice, but they also allow for positivity.

Theorem 1 shows that, with positivity, the probability of choosing $x \in B$ must satisfy

$$\Pr(x|B) = \frac{\exp\{\lambda \cdot u_x + \kappa \cdot \phi_x\}}{\sum_{x' \in B} \exp\{\lambda \cdot u_{x'} + \kappa \cdot \phi_{x'}\}}, \tag{1}$$

for some $\lambda, \kappa \in \mathbb{R}$. I call this model focal choice adjusted logit (Focal). The representation is unique under a standard richness assumption and can be inverted (Theorem 2) to infer u and ϕ up to affine transformation from observed choice probabilities. That is, u and ϕ are distinguishable if they are not collinear, e.g. if they are functions of option attributes that can be varied independently. In studies of social preferences, such independence follows from the standard assumption that utility is a function of option payoffs and presentation affects choice via payoff independent option attributes such as positioning and labeling.[1]

[1] In the above dictator game, varying the number of tokens B while holding the total value of all tokens constant implies changes in the payoffs (and thus utilities) associated with options, without affecting the "roundedness" of options (number of tokens transferred) as such. Similarly, subsidizing or taxing transfers

Both results, Theorems 1 and 2, are surprising from an ex-ante perspective. Theorem 1 shows that structural models of stochastic choice can be defined without making functional form assumptions on error distributions, which has been disputed by critiques of structural work. For example, in response to such critiques, Rust (2014, p. 820) concedes that "there is an identification problem that makes it impossible to decide between competing theories without imposing ad hoc auxiliary assumptions (such as parametric functional form assumptions)". Focal relies solely on positivity and four invariance statements, which suffices thanks to a novel analytical approach via functional equations. Theorem 2 shows that despite their linearly additive relationship, u and ϕ are both identified under a simple independence assumption satisfied by received models of social preferences. That is, utility and relative focality are identified without assumptions restricting their functional forms, which enables nonparametric studies of utility and stochastic choice. Finally, based on Gul and Pesendorfer (2001), Focal illustrates the applicability of this celebrated choice model in econometric analyses and nests a number of specific thought-process models allowing for "confounds" besides utility, most notably Mattsson and Weibull (2002) and Matejka and McKay (2015). This suggests that Focal indeed provides an adequately general characterization of choice in the presence of presentation effects.

In Sections 4 and 5, the Focal model is tested. The general idea underlying this test is to determine whether Focal can be a useful model in applied work—on data sets as they are typically analyzed under econometric assumptions as they are typically made. The questions asked address the main concerns in applied work: Does Focal enable more accurate counterfactual predictions than existing models? Does it provide consistent and reliable estimates across studies? Do other models fail in this respect, and is the relevance of presentation effects as substantial as suggested by Figure 1? This approach, testing structural models by evaluating validity and consistency in-sample and out-of-sample using typical data sets and standard specification, follows Keane and Wolpin (2007) and Keane (2010a,b).[2] Other analyses of behavioral models evaluating out-of-sample predictions include analyses of decision under risk (Harless and Camerer, 1994; Wilcox, 2008; Hey et al., 2010), learning (Camerer and Ho, 1999), strategic choice in normal-form games (Camerer et al., 2004), and social preferences (De Bruyn and Bolton, 2008).

Counterfactual predictions are required for policy recommendations and to form hypotheses for experimental studies. To predict, the analyst needs to factor out presentation in the original context and to factor in presentation in the context to be predicted. Given the (small) size of experimental data sets, this may well fail even if the choice model is sound. Consistency of estimates requires analysts to control for presentation effects in experimental work, which necessitates additional information and thus may well fail given current experimental designs. Finally, Focal may be of significant help in applied work,

changes payoffs and utilities without affecting roundedness of options.

[2]Implicitly, I test the model in its entirety and ask whether it does what it is supposed to do in the context where it is supposed to be applied. An alternative approach is to test the axioms or assumptions in isolation, one-by-one in dedicated, context-free experiments. This approach is possible here as much as it is possible with other models, but it is somewhat inadequate in the present analysis. My basic premises are that behavior is presentation dependent in experiments designed by analysts. Hence, the results of dedicated analyses of axioms in isolated, artificial contexts are not informative. A case in point are tests of IIA, as discussed in Section 3.2. Briefly, it is straightforward to falsify IIA, by duplicating options, but this is not a concern in experimental studies, as experiments are designed by researchers diligently avoiding duplicate options.

using "significant" in its statistical sense, without being of economic relevance. In this case, Focal might not be worth the additional effort required, albeit small.

The plots in Figure 1 may help putting such concerns into context. The presentation effects are drastic, recall Figures 1a and 1c for example. They are also economically highly relevant as they affect the level of transfers and the sign of comparative statics. Perhaps most strikingly, while taxation and subsidization of transfers may well prove irrelevant, as in HJ06 and FKM07, changing the presentation can increase transfers by 50%, from 20% of endowments (FKM07) to 30% of endowments (HJ06). Considering this scale of the phenomena related to presentation, it reasonable to hypothesize that addressing presentation is highly important already in typically sized and structured experiments.

The econometric results of re-analyzing dictator experiments strongly corroborate this hypothesis. In sample, Focal explains around 85% of observed variance across dictator experiments (in terms of the pseudo-R^2), while benchmark models such as logit explain around 60% of observed variance. Out-of-sample, i.e. making counterfactual predictions, Focal captures around 80% of observed variance, while benchmark models capture around 40%. Thus, Focal's predictions are much more accurate, and in this respect Focal marks an improvement of substantial proportion. Next, the difference between Focal's numbers, 85% and 80%, is insignificant, showing that Focal's estimates are indeed consistent across data sets and that its predictions (almost) have in-sample accuracy. In turn, the differences for the benchmark models, roughly 60% in-sample to 40% out-of-sample, are highly significant and show that existing models yield significantly inconsistent estimates.

Finally, to analyze relevance, I relate the improvement in model adequacy gained by controlling for presentation (using Focal) to the adequacy gained by controlling for subject heterogeneity. The baseline model is the representative agent logit model, controlling for neither heterogeneity nor presentation. Its premise is that the average degree of altruism in a population is informative, which captures about 20% of observed variance in-sample and about -5% out-of-sample (i.e. "minus five per cent", it predicts worse than a model predicting uniform randomization)—its predictions are slightly worse than useless and the existence of a representative agent has to be rejected. In relation to this baseline, controlling for heterogeneity improves adequacy by around 40 percentage points, both in-sample and out-of-sample. Controlling for presentation adds another 30 percentage points in-sample and 40 percentage points out-of-sample. That is, controlling for presentation is orthogonal to and of the same importance as controlling for heterogeneity—in typical experimental data sets. Hence, experimental studies need to control for both, heterogeneity and presentation. There is no reason for doing one without the other, and future experimental designs can facilitate controlling for presentation, similarly to the efforts made to control for heterogeneity in current designs and analyses. These and related implications are discussed in the concluding Section 6.

3 Focal choice adjusted logit

Decision maker DM chooses an option $x \in B$ from a finite budget set $B \subset X$. Based on Gul and Pesendorfer (2001), we allow DM's decision to be a function of utility $u : X \to \mathbb{R}$ and focality $\phi : X \to \mathbb{R}$ of options. Utility u_x represents the utility that DM derives from option $x \in X$ ("commitment ranking" in the words of Gul and Pesendorfer), and focality ϕ_x represents a secondary choice factor induced by the presentation of options ("temptation ranking" in Gul and Pesendorfer). The pair (u,ϕ) is called "context" of DM's choice task. The set of contexts (u,ϕ) is denoted as $\mathcal{C}$, and the set of possible budgets B is the set of finite subsets of X, denoted as $P(X)$. Given context $(u,\phi) \in \mathcal{C}$, the probability that DM chooses $x \in B$ from budget $B \in P(X)$ is denoted as $\Pr(x|u,\phi,B)$.

One may think of u as a function of option payoffs and ϕ as a function of option labels or positioning, which in principle are independent of another and could be conditioned on directly—at the cost of maintaining otherwise unnecessary notation. Conditioning on the unknown (u,ϕ) in $\Pr(x|u,\phi,B)$ is a notational shortcut, but it does not assume that u or ϕ are known or directly observable. To the contrary, the objective is to infer u and ϕ from choice, and to this purpose, we first analyze how to represent choice probabilities in terms of u and ϕ (Section 3.1) and then invert the system to show that both u and ϕ are identified up to affine transformation based on simple experiments (Section 3.3). Section 3.2 discusses assumptions and axioms underlying the representation result.

3.1 Representation

Assumption 1 characterizes the analytical framework. First, for any given context (u,ϕ) available to an experimenter, all affine transformations of this context (i.e. of u and ϕ) are equally possible—which will allow me to show that affine transformations actually are indistinguishable for the experimenter in the sense that u and ϕ are identified only up to affine transformation.[3] In addition, this assumption ensures that u and ϕ can be varied independently, which is required for identification. Second, I exclude the trivial cases that utility and focality are collinear and that choice probabilities are constant. Third, and this guarantees uniqueness, the set of possible options X as well as the images of u and ϕ are convex subsets of $\mathbb{R}$. Convexity of X obtains in standard examples, as say lottery probabilities and dictator transfers can be varied continuously although DM's option set is a finite set $B \subset X$ constructed by the experimenter. Convexity of the images then follows from standard continuity assumptions. Similar richness assumptions are made in most studies analyzing foundations of choice, see e.g. Gul et al. (2014) and Fudenberg et al. (2015) for examples in the context of stochastic choice.

Assumption 1. For any $(u,\phi) \in \mathcal{C}$, using $u[X] = \{u_x|x \in X\}$ and $\phi[X] = \{\phi_x|x \in X\}$,

1. *Choice tasks:* $(a_u + b_u \cdot u, a_\phi + b_\phi \cdot \phi) \in \mathcal{C}$ for all $(a_u,b_u),(a_\phi,b_\phi) \in \mathbb{R}^2$.
2. *Non-triviality:* The functions u and ϕ are not collinear, and there exist $x,x' \in B \in P(X)$ such that $\Pr(x|u,\phi,B) \neq \Pr(x'|u,\phi,B)$.
3. *Convexity:* $X, u[X], \phi[X]$ are convex, non-singleton subsets of $\mathbb{R}$.

[3] With slight abuse of notation, I will identify all real numbers as constant functions such that addition and multiplication of a function with a real are well-defined. Thus, for any $u : X \to \mathbb{R}$ and any $a,b \in \mathbb{R}$, $u' = a + bu$ is equivalent to $u'_x = a + bu_x$ for all $x \in X$. Note also that I allow for negative affine transformations (i.e. $b < 0$). For, based on choice data alone, we are not able to distinguish a DM maximizing utility from a DM minimizing utility without assuming monotonicity. Thus, without monotonicity, utility and focality are not identified up to positive affine transformation. See Benkert and Netzer (2015) for further discussion.

Choice is assumed to comply with five axioms: positivity, IIA, invariance to translation of utility u and focality ϕ, and equality of utility and focality between two options implies equality of relative choice probabilities (in relation to third options). The latter implies that u and ϕ are the only systematic factors of choice. A related but stronger assumption is to require that choice probabilities are weakly monotone in utility and focality. This implies "completeness" but not vice versa. Weak monotonicity is not required.

Behavioral axioms For all $(u,\phi) \in \mathcal{C}$, all $B \in P(X)$, all $x \in B$ and all $r \in \mathbb{R}$,

1. *Positivity:* $$\Pr(x|u,\phi,B) > 0,$$
2. *IIA:* for all $B' \in P(X)$ and all $x,y \in B \cap B'$, $$\frac{\Pr(x|u,\phi,B)}{\Pr(y|u,\phi,B)} = \frac{\Pr(x|u,\phi,B')}{\Pr(y|u,\phi,B')},$$
3. *Narrow bracketing:* $$\Pr(x|u,\phi,B) = \Pr(x|u+r,\phi,B),$$
4. *Relative focality:* $$\Pr(x|u,\phi,B) = \Pr(x|u,\phi+r,B),$$
5. *Completeness:* for all $(u,\phi),(u',\phi') \in \mathcal{C}$, $B = \{x,y\}$, and $B' = \{x',y'\}$, $$(u_x,\phi_x) = (u'_{x'},\phi'_{x'}) \text{ and } (u_y,\phi_y) = (u'_{y'},\phi'_{y'}) \quad \Rightarrow \quad \frac{\Pr(x|u,\phi,B)}{\Pr(y|u,\phi,B)} = \frac{\Pr(x'|u',\phi',B')}{\Pr(y'|u',\phi',B')}.$$

These assumptions uniquely pin down the functional form of choice probabilities and imply that choice probabilities have a focal choice adjusted logit (Focal) representation. The detailed discussion follows the presentation of the formal result.

Definition 1 (Focal choice adjusted logit (Focal))**.** DM's choice profile Pr has a Focal representation if there exist $\lambda,\kappa \in \mathbb{R}$ such that

$$\Pr(x|u,\phi,B) = \frac{\exp\{\lambda \cdot u_x + \kappa \cdot \phi_x\}}{\sum_{x' \in B} \exp\{\lambda \cdot u_{x'} + \kappa \cdot \phi_{x'}\}} \qquad \text{for all } x \in B \in P(X) \text{ and } (u,\phi) \in \mathcal{C}. \quad (2)$$

Theorem 1. *Given Assumption 1, the following two statements are equivalent.*

1. *Choice profile* Pr *satisfies Axioms 1–5*
2. *Choice profile* Pr *has a Focal representation and* $(\lambda,\kappa) \in \mathbb{R}^2$ *are unique*

Intuition Positivity and IIA imply that choice probabilities are functions of "choice propensities" (Luce, 1959). Without further information, the choice propensity of $x \in X$ is defined only in relation to a benchmark option $y \in X$ (McFadden, 1974), and as IIA applies to each context (u,ϕ) separately, propensities may be context dependent, implying

$$\Pr(x|u,\phi,B) = \frac{V(x,y\,|\,u,\phi)}{\sum_{x' \in B} V(x',y\,|\,u,\phi)} \qquad \text{with} \qquad V(x,y\,|\,u,\phi) = f_{u,\phi}(u_x,u_y,\phi_x,\phi_y,x,y).$$

Any $y \in X$ may serve as benchmark option, and by convexity of X, the reference to the benchmark can be dropped, i.e. $V(x,y\,|\,u,\phi) = \tilde{f}_{u,\phi}(u_x,\phi_x,x)$ for some function $\tilde{f}$. By Axiom 5, completeness, options that are equivalent in terms of both utility and focality have equal choice propensities in any choice task, implying that choice propensities can be represented independently of x. Formally, a family of functions $\tilde{V}_{u,\phi}$ exists such that $\tilde{V}_{u,\phi}(u_x,\phi_x) = \tilde{f}_{u,\phi}(u_x,\phi_x,x)$ for all $x \in X$, and all contexts (u,ϕ). Given this characterization of propensities, narrow bracketing and relative focality imply

$$\frac{\tilde{V}_{u,\phi}(u_x,\phi_x)}{\tilde{V}_{u,\phi}(u_{x'},\phi_{x'})} = \frac{\tilde{V}_{u+r_u,\phi+r_\phi}(u_x+r_u,\phi_x+r_\phi)}{\tilde{V}_{u+r_u,\phi+r_\phi}(u_{x'}+r_u,\phi_{x'}+r_\phi)} \quad \text{for all } x,x' \in X \text{ and } r_u, r_\phi \in \mathbb{R}, \tag{3}$$

which in turn implies $\tilde{V}_{u+r_u,\phi+r_\phi}(u_x+r_u,\phi_x+r_\phi) = \tilde{V}_{u,\phi}(u_x,\phi_x)\cdot g(r_u,r_\phi)$ for some function g. Due to the context dependence of $\tilde{V}$, $\tilde{V}_{u+r_u,\phi+r_\phi} \neq \tilde{V}_{u,\phi}$ is possible, but "completeness" restricts context dependence by allowing the functional equation to be expressed as $h(u_x+r_u,\phi_x+r_\phi) = h(u_x,\phi_x)\cdot g(r_u,r_\phi)$ for some function $h:\mathbb{R}^2 \to \mathbb{R}$. The main technical difficulty is that h is not necessarily differentiable.[4] By positivity, the logarithmic transformation $\tilde{h} = \log h$ and $\tilde{g} = \log g$ is admissible, which yields the Pexider functional equation $\tilde{h}(u_x+r_u,\phi_x+r_\phi) = \tilde{h}(u_x,\phi_x) + \tilde{g}(r_u,r_\phi)$, and by their relation to probabilities, $\tilde{h}$ is bounded from above for all values in the images of u and ϕ (in any context),[5] which each have positive length by "convexity" (Assumption 1). This implies that all solutions of $\tilde{h}$ are linear in u_x and ϕ_x, and the non-collinearity of u and ϕ implies that the respective coefficients (λ,κ) are unique. Thus, $\tilde{h}(u_x,\phi_x) = \lambda u_x + \kappa\phi_x + c_x$ for all x, with unique $\lambda,\kappa \in \mathbb{R}$ and $c: X \to \mathbb{R}$ (Aczél and Dhombres, 1989). Using $\tilde{V}_{u,\phi} = \exp\tilde{h}$,

$$\Pr(x|u,\phi,B) = \frac{\exp\{\lambda\cdot u_x + \kappa\cdot\phi_x + c_x\}}{\sum_{x'\in B}\exp\{\lambda\cdot u_{x'} + \kappa\cdot\phi_{x'} + c_{x'}\}}. \tag{4}$$

Completeness implies that c_x is constant in x and cancels out, yielding the Focal representation. The detailed proof is relegated to the appendix.

3.2 Discussion

Foundation for u and ϕ Gul and Pesendorfer (2001) introduce the notion of set-based preferences to capture decision makers experiencing temptation and exerting self-control. The basic model is general, essentially capturing a DM experiencing arbitrary "costs" when having to choose between options. These costs could be due to temptation, as in Gul and Pesendorfer's leading example, due to the relative focality of options as a result of their presentation, as in the examples considered here, or due to mechanical or computational difficulties in implementing or evaluating options, as in the thought-process models discussed below. Gul and Pesendorfer's model encompasses this range of applications thanks to its general formulation, extending the standard framework of a DM with preferences over outcomes "simply" by allowing that DM's evaluation of a set of options differs from the evaluation of the best option in the set. This difference reflects the costs of choosing from the respective set of options, it depends on the available alternatives, and may

[4] For purpose of illustration, consider the simple case that propensities $\tilde{V}$ are context independent, differentiable in u_x and independent of ϕ_x. Then we obtain $\tilde{V}(u+r) = \tilde{V}(u)\cdot g(r)$ and after differentiating with respect to r, at $r=0$ we obtain $\tilde{V}'(u) = \tilde{V}(u)\cdot g'(0)$. The solution of this differential equation is $\tilde{V}(u) = \exp\{\lambda u + c\}$ with $\lambda = g'(0)$ and $c \in \mathbb{R}$.

[5] Briefly, the choice probabilities $\Pr(x|u,\phi,B)$ are bounded from above at 1, the propensities $\tilde{V}(u_x,\phi_x)$ are therefore bounded from above at some positive real number, finally so are log-propensities $\tilde{f} = \log\tilde{V}$.

result from temptation, focality, or any other distraction. Gul and Pesendorfer additionally assume that the evaluation of a set is not worse than the evaluation of the worst option in this set, which in conjunction with completeness, transitivity, continuity and independence implies that functions $u, v : X \to \mathbb{R}$ exist such that DM's choice can be represented as one maximizing $u_x + v_x$ in stage 2 of their model (when facing "temptation"). This result is the foundation of the model proposed here, with the notational adaptation that we focus on relative focality ϕ in light of the applications we have in mind.

Normative appeal of the axioms If DM maximizes $u_x + v_x$ across available options, or $u_x + \phi_x$ in our notation, then choice satisfies IIA, is invariant to translation of u and ϕ, and behavior is described "completely" by u and ϕ (in the sense of Axiom 5). In turn, if choice satisfies these axioms, then it can be represented as maximization of $u_x + \phi_x$. To be precise, if $u, \phi : X \to \mathbb{R}$ are exogenously given, then these axioms imply that DM maximizes $\alpha u_x + \beta \phi_x$ for some $\alpha, \beta \in \mathbb{R}$, which reflects that u, ϕ are identified only up to affine transformation. That is, Axioms 2–5 are, aside from the fact that they allow for stochastic choice, equivalent to Gul and Pesendorfer's representation of choice in "stage 2" of their model, and in this sense, the axioms used here use Gul and Pesendorfer's result as foundation, represent them as invariance statements about choice probabilities (Axioms 2–5) and add positivity (Axiom 1) to inject stochastic choice. From this perspective, the axioms are normatively adequate given the results of Gul and Pesendorfer (2001).

Nature and testability of Axioms 3 and 4 Axioms 3 and 4 relate the observable choice profile Pr to unobservable entities u and ϕ. This is standard practice in modeling stochastic choice (though usually less explicit), as any model of stochastic choice links Pr and u based on some explicit or implicit assumption. To clarify, the existing literature distinguishes two approaches toward axiomatic foundations of logit and generalizations thereof. The first approach may be called unconditional logit and follows Luce (1959). Basically, positivity and IIA imply that choice probabilities satisfy

$$\Pr(x|B) = \frac{\exp\{u_x\}}{\sum_{x' \in B} \exp\{u_{x'}\}}$$

for some function $u : X \to \mathbb{R}$. Function u can be intuitively interpreted as stimulus or utility. Given IIA and positivity, choice probabilities also satisfy

$$\Pr(x|B) = \frac{\tilde{u}_x}{\sum_{x' \in B} \tilde{u}_{x'}}$$

for some function $\tilde{u} : X \to \mathbb{R}$. If one assumes that either u or $\tilde{u}$ represents utility up to affine transformation, one implicitly assumes that either of these functional forms is adequate and thus makes an assumption linking choice probabilities and (unobservable) utility.

The second approach is called conditional logit and follows McFadden (1974). Axiom 3 of McFadden explicitly links choice probabilities and unobservable utility, and in addition, it is equivalent to assuming that binomial choice is binomial logit (Breitmoser, 2016). Furthermore, both conditional and unconditional logit characterizations make specific functional form assumptions while relating probabilities and utilities, functional form assumptions that are not behaviorally founded and frequently criticized for this reason (for discussion, see Rust, 2014). In contrast, the narrow bracketing axiom used here is a substantially weaker invariance assumption and implied by Gul and Pesendorfer's results—which resolves the canonical critique of structural models.

Regarding testability, since all econometrically applicable models of stochastic choice require assumptions or axioms linking choice probabilities and (unknown) utilities, direct tests by manipulating utilities are difficult to conceive, in all cases. Indirect tests are straightforward, however. For example, based on the identification result provided below, logarithmized choice probabilities are affine functions of utilities and focalities, which can be used to test via regression if the linear functional form indeed provides a good fit. If not, translation invariance with respect to utility or focality is violated. Alternatively, we can test if the model as a whole allows to capture and predict behavior.[6] This approach allows to detect inadequate assumptions by relating the model's adequacy to other models based on weaker assumptions, and it allows to evaluate model adequacy in domains of empirical relevance. Since behavior may be domain specific, this approach makes for a particularly informative test of models from the perspective of applied work and is adopted here.

Relation to thought-process models A number of studies analyze models of the thought process underlying choice and then derive conditions for this process to generate say logit choice. The perhaps best-known such model is the random utility model (Thurstone, 1927) where the utilities of all options are randomly perturbed prior to DM's choice. Logit obtains if the utility perturbation have extreme value type 1 distribution (McFadden, 1974). Mattsson and Weibull (2002) consider a DM facing costs implementing specific options with certainty, say bowling a strike. A rational DM trades off precision and costs, therefore ends up choosing stochastically, and under certain assumptions on the cost function, choice satisfies the following generalized form of logit.

$$\Pr(x|B) = \frac{\exp\{\lambda u_x + w_x\}}{\sum_{x' \in B} \exp\{\lambda u_{x'} + w_{x'}\}}$$

The additional term w_x reflects the prior choice probabilities, a default choice pattern resulting if DM does not attempt to exert any control. Matejka and McKay (2015) obtain the same generalized logit model assuming DM has to spend time to study the utilities associated with the available options and is "rationally inattentive". Again, DM maximizes expected utility of the choice less the costs of studying utilities, and if costs relate to the Shannon entropy, the above generalized logit model results, again with w_x representing the prior choice probabilities that obtain if DM does not study utility at all.

Both of these models fit the Gul-Pesendorfer framework, most notably set-betweenness, and thus are contained as special cases. As a result, they also are contained as special cases in Focal, which shows as Focal allows for an additional free parameter κ. This

[6] For example, Keane (2010b, p. 15) argues that the usefulness of structural models be determined resting on two questions: "(1) Does the model do a reasonable job of fitting important dimensions of the historical data on which it was fit? (2) Does the model do a reasonable job at out-of-sample prediction – especially when used to predict the impact of policy changes that alter the economic environment in some fundamental way from that which generated the data used in estimation?" Instead of policy changes, I consider changes in the way the choice tasks are presented to experimental subjects.

generalization reflects that Focal avoids specific assumption about the thought process and cost functions or entropies, but considering that avoiding these assumptions merely costs an additional parameter suggests that Focal's representation result is surprisingly tight.

Positive appeal of the axioms Positivity implies that choice is stochastic, i.e. that DM does not manage to consistently maximize utility. This assumption is a widely adopted, parsimonious explanation for inconsistencies observed in behavioral analyses, as discussed by Hey (1995) and Wilcox (2008) in detail.[7] McFadden (1974) also notes that positivity is empirically indistinguishable from rational choice, as the probabilities may be arbitrarily close to zero, and relatedly, all structural analyses (including least squares) allow for stochastic choice and assume positivity.

IIA assumes that choice does not exhibit similarity effects, the limitation of which was first clarified by Debreu (1960)'s red-bus/blue-bus example. The Focal model can be generalized straightforwardly when similarity effects are present, but arguably such generalizations are not necessary in utility and demand estimation. In general, similarity effects can be captured by nested logit, a random utility model with perturbations following a generalized extreme value (GEV) distribution. Focal also is a random utility model and rewriting Eq. (2), Focal can be represented as

$$\Pr(x|u,\phi,B) = \frac{\exp\{\lambda \cdot \left(u_x + \tilde{\kappa} \cdot \phi_x\right)\}}{\sum_{x' \in B} \exp\{\lambda \cdot \left(u_{x'} + \tilde{\kappa} \cdot \phi_{x'}\right)\}} \qquad \text{for all } x \in B \in P(X).$$

Hence, Focal admits GEV perturbations and thus nested (focal choice adjusted) logit. Nested logit models are used in transportation research, but rarely in empirical IO and to my knowledge never in experimental analyses. The reasons appear to be that demand functions explicitly capture similarity effects (Nevo, 2000) and that laboratory experiments are generally designed to avoid similar (redundant) options for clarity (Davis and Holt, 1993). Least squares analyses also obey IIA. Thus, in the context of utility and demand estimation, IIA is not considered "problematic" in existing work, but as indicated, generalizations are possible and I will consider a model relaxing IIA as a benchmark below.

Narrow bracketing in general refers to the phenomenon that "background utility" is behaviorally irrelevant (Read et al., 1999). Under rational choice, this obtains if DM's utility is the sum of "decision utility" from the task being analyzed and "background utility" from experiences outside of this task, including earlier tasks in the experiment. Such invariance with respect to variation in background utility is generally assumed in behavioral analyses, including least squares analyses, in the sense that background utility is not modeled or analyzed (Harrison et al., 2007). It is consistent with the robust finding that behavior in experiments is independent of socio-economic background variables such as income or wealth (Gächter et al., 2004; Bellemare et al., 2008, 2011).

Relative focality requires that choice is invariant to translating focality. As illustration, consider the choice of an integer and assume that multiples of 100 have focality 2,

[7] Experimental subjects act "inconsistently" in the sense that they choose dominated options (Birnbaum and Navarrete, 1998; Costa-Gomes et al., 2001), violate revealed preference (Andreoni and Miller, 2002; Fisman et al., 2007), and choose inconsistently even after controlling for wealth and portfolio effects (Camerer, 1989; Starmer and Sugden, 1991).

other multiples of 10 have focality 1, and other integers have focality 0. Here, plain integers are assumed to have the base level of focality (zero), but it appears to be equally plausible to state that multiples of 10 have the basic "zero" focality or the multiples of 100, provided all other focalities are translated accordingly. This intuition holds if "relative focality" is relevant, i.e. the differences in focalities between options and not their absolute values. Axiom 4 reflects this intuition, and the results below suggests that it is adequate in the sense that choice patterns will turn out predictable across very different option sets.[8]

Finally, "completeness" implies that utility and focality are the only systematic factors of choice, reflecting the scope of the present analysis and model.

3.3 Identification

The representation result derives the choice profile Pr if utility and focality are given and implies that utility and focality are additively separated—suggesting they may be indistinguishable. The following result shows that u and ϕ are identified without additional assumptions from choice probabilities observed in adequate experiments.

Theorem 2. *Fix* $(u,\phi) \in \mathcal{C}$*, and assume* Pr *has a Focal representation, i.e. Axioms 1–5 and Assumption 1. The choice utility* $v_x := \log \Pr(x|u,\phi,X)$*, for all* $x \in X$*, satisfies*

$$v_x = a + \lambda\, u_x + \kappa\, \phi_x \tag{5}$$

with (λ,κ) *as in the Focal representation and* $a = \inf_x v_x - \inf_x(\lambda u_x + \kappa\phi_x)$*. This implies that, for any* $B \subset X$*, observations from two (adequate) choice tasks suffice to identify both,* $u : B \to \mathbb{R}$ *and* $\phi : B \to \mathbb{R}$*, up to affine transformation.*

For intuition, assume that utility is a mapping from option payoffs to the reals, while focality is a mapping from option labels to the reals. When we permute option payoffs without changing option labels, Eq. (5) implies that identification of u and ϕ requires observations from only two choice tasks and is possible without functional form assumptions (i.e. non-parametrically). This is shown constructively, in order to clarify that there are no fundamental limitations to identification implied by Focal. Let me note, however, that the nonparametric approach may be inefficient if the analyst has prior knowledge about say utility functions, seeks to test utility theories, or wishes to estimate say a degree of altruism, to pool observations from different games (which requires a general notation of utility), or to predict counterfactuals.

Task 1 implements a payoff function $\pi : B \to \mathbb{R}^n$ on the budget set $B = \{0,\dots,10\}$; let $\underline{x} = 0$ and $\overline{x} = 10$. Task 2 maintains budget set and option labels, but permutes the mapping from options to payoffs, for example by rotation toward $\tilde{\pi} : B \to \mathbb{R}^n$ as in

$$\tilde{\pi}(x) = \pi(x-1) \text{ for all } x > \underline{x} \qquad \text{and} \qquad \tilde{\pi}(\underline{x}) = \pi(\overline{x}).$$

[8] I am not aware of explicit evidence, but relativity of focality can be tested straightforwardly. If DM first has to choose an integer between 1 and 30, and secondly a multiple of 10 between 10 and 300, reassigning utilities appropriately, relativity predicts that the respective choice patterns should be equal.

Since utility is a mapping from option payoffs to the reals, this implies $\tilde{u}_x = u_{x-1}$ and $\tilde{u}_{\underline{x}} = u_{\overline{x}}$. Let $\Pr \in \Delta(X)$ and $\tilde{\Pr} \in \Delta(X)$ denote the choice probabilities in the two tasks. By Theorem 2, $\log \Pr(x) = a + \lambda u_x + \kappa\phi_x$ and $\log \tilde{\Pr}(x) = \tilde{a} + \lambda \tilde{u}_x + \kappa\phi_x$, implying

$$\begin{aligned} \log \tilde{\Pr}(x) - \log \Pr(x-1) &= \tilde{a} - a + \kappa\phi_x - \kappa\phi_{x-1} \qquad \forall x > \underline{x}, \qquad \text{and} \\ \log \tilde{\Pr}(\underline{x}) - \log \Pr(\overline{x}) &= \tilde{a} - a + \kappa\phi_{\underline{x}} - \kappa\phi_{\overline{x}}, \end{aligned}$$

with $\tilde{a} - a$ constant. Letting $\phi_{\underline{x}} = 0$, focality $\phi : B \to \mathbb{R}$ is thus identified up to affine transformation.[9] Given this, and using $a + \lambda u_x = \log \Pr(x) - \kappa\phi_x$, $u : B \to \mathbb{R}$ is also defined up to affine transformation. Lemma 1 in the appendix establishes this result formally, and Theorem 2 ensures that the estimates are approximately consistent.

A few more points may be worth noting about identification. First, the identification u and ϕ is entirely nonparametric, i.e. their separation is achieved without making identifying assumptions or assuming specific functional forms to structure noise, utility, or presentation effects (focality). Second, rotation of the payoff function ensures identification up to affine transformation using observations from two tasks, but adding further tasks does not refine identification beyond affine transformation (obviously, extending sample size improves finite sample properties). Third, rotation is far from being the only scheme enabling identification, but not every permutation of options does. The best-known inadequate permutation is the reversion of option lists in order to nullify ordering effects. This approach would be adequate if we knew that focality is linear in list positioning, but evidence suggests that focality tends to be nonlinear and potentially u-shaped (see also Rubinstein and Salant, 2006, and Feenberg et al., 2017). For example, if both first and last option have high focality and all other options have low focality, then reversion has no effect. Acknowledging this, an equation system similar to above has to be solved for identification, but the equation system resulting from order reversion is generally dependent, implying that identification by reversion is impossible (see Lemma 2 in the appendix).

Finally, utility can be estimated also semi-parametrically, by determining $\kappa\phi_x$ nonparametrically (as above) and regressing the resulting "raw utilities" $a + \lambda u_x = \log \Pr(x) - \kappa\phi_x$ on payoff profiles, or parametrically by adopting functional forms of both u and ϕ. Their parameters as well as λ and κ can be estimated by (non-linear) regression of choice utility v on payoffs and labels under the standard identification requirements. The main requirement is that log-propensities at the true parameters are not translations of log-propensities at alternative parameter vectors (see e.g. Greene, 2003). This is satisfied if both utility and focality are choice relevant ($\lambda, \kappa > 0$), no two parameterizations of u (or ϕ) are affine transformations of one another, and utility and focality at the true parameters are not collinear. These conditions are reasonable and can be checked straightforwardly. Since the additive constant a and the scaling factors λ, κ are free parameters, identification of u and ϕ is again unique up to affine transformation. I will estimate parameters by maximum likelihood, which are identified under the same conditions as regression.

[9] The preceding set of equations is a system of $\overline{x} - \underline{x} + 1 = 11$ linear equations with 11 unknowns: all $\kappa\phi_x$ with $x > \underline{x}$ and a constant $c := \tilde{a} - a$. Since κ and ϕ_x are not separately identified and $\phi_{\underline{x}} = 0$ was set, identification is unique up to affine transformation. Lemma 1 in the appendix provides the details.

Table 1: The data sets

	#Treatments	#Options	#Observations	Transfer ratios
"Manual" dictator games				
AM02 (Andreoni and Miller, 2002)	8	41–101	176×8	$3:1,\ldots,1:3$
HJ06 (Harrison and Johnson, 2006)	10	41–101	57×10	$1:1,\ldots,1:4$
CHST07 (Cappelen et al., 2007)	6	401–1601	96×2	1:1
"Graphical" dictator games				
FKM07 (Fisman et al., 2007)	50	500–1000	76×50	$4:1,\ldots,1:3$

4 Testing Focal: Data and general approach

The purpose of the econometric test is to apply the model in conditions typical for applied work, i.e. on standard data sets making standard assumptions. This section explains why dictator experiments are selected for this purpose and states the typical assumptions on utility, subject heterogeneity, and focality in this context, where the latter is borrowed from the statistics literature on round-number effects in survey responses.

The selected data sets are from experiments on dictator games. Dictator games enable an analysis of utility and focality in a context where utility and focality individually appear to be understood and additional factors can be ruled out. Specifically, there is consensus that utilities u are adequately captured by CES functions, there are unambiguous presentation effects (round-number effects), and the choice task does not involve confounds due to e.g. risk, probability weighting, or (strategic) uncertainty. There does not seem to be a better-suited class of experiments for our purpose. Further, there exist four well-known experimental analyses designed to estimate utility functions and differing only in presentational aspects, which appears to be uniquely true for dictator games. The four experiments elicit multiple choices per subject, which allows to disentangle utility and noise, but otherwise they are representative for hundreds of small-scale dictator experiments (Engel, 2011). Finally, one of these experiments (Fisman et al., 2007) avoids presentation effects due to round numbers by using a graphical user interface, which provides "counterfactual" information when analyzing consistency and reliability of estimates across studies.

Definition 2 (Generalized dictator game)**.** DM chooses an option $x \in \{0,1,\ldots,B\}$. Given x, the dictator's payoff is $\pi_1(x) = \tau_1 \cdot (B-x)$ and the recipient's payoff is $\pi_2 = \tau_2 \cdot x$.

Table 1 provides an overview of the data sets and Figure 1 above provides selected histograms of observed choices (the supplement provides all histograms). In conjunction, these studies provide a fairly comprehensive picture of dictator choice. Fisman et al. (2007, FKM07) use the graphical user interface preventing round-number effects, while the other studies elicit choices manually, implying round-number patterns. Cappelen et al. (2007, CHST07) allow for budgets up to $B = 1600$, and subjects primarily choose multiples of 100, Andreoni and Miller (2002, AM02) and Harrison and Johnson (2006, HJ06) allow for budgets up to $B = 100$ and choices mainly are multiples of 10. In AM02, the Leontief (payoff-equalizing) choice is generally a multiple of 10 or 25, but in HJ06, it is often a plain integer. The latter drastically affects the relative frequency of the Leontief choice, showing that the Leontief choice is frequent if and only if it is a round number.

Table 2: Distribution of choices across “round” numbers

	Percentage of choices with greatest factor …												
Experiment	0	1000	500	250	100	50	25	10	5	2.5	1	0.5	0.1
AM02	**39**				1	9	7	**33**	6	0	4		
HJ06	**22**				3	4	7	**39**	14	0	10		
FKM07	**25**				0	0	0	0	2	1	4	5	**63**
CHST07	**30**	0	5	1	**62**	1	0	0	0	0	0		

Note: For each experiment, these percentages are pooled (and averaged) across treatments. The numbers do not always add up to exactly 100 due to rounding errors.

Table 3: Level of focality depending on the “roundedness” of option x

Greatest factor of x	1000	500	100	50	10	5	1	0.5	0.1
Level of focality ϕ_x	6	5	4	3	2	1	0	-1	-2

Note: For example, the option to transfer $x = 60$ tokens has the greatest factor 10 in the above list and hence focality $\phi_x = 2$. In addition, $\phi_0 := 2$. The base level of “zero focality” can be shifted arbitrarily.

Utility Following Andreoni and Miller (2002), most dictator game analyses estimate utility functions exhibiting constant elasticity of substitution (CES) between dictator income π_1 and recipient income π_2. That is, DM’s utility is

$$u(\pi_1, \pi_2) = \big((1-\alpha)\cdot(1+\pi_1)^\beta + \alpha\cdot(1+\pi_2)^\beta\big)^{1/\beta}. \tag{6}$$

Here, α represents the degree of altruism and β represents the degree of efficiency concerns. Subjects are efficiency concerned with $\beta = 1$ and equity concerned as $\beta \to -\infty$.

Focality Table 2 reviews the numbers chosen in these experiments. In the experiments with manual choice, subjects rarely choose plain integers. Mostly, they choose multiples of 10 and 100, as described above. In addition, subjects tend to choose multiples of 5 and 50 reasonably regularly, while multiples of 2.5 and 25 are chosen a little less frequently. These choice patterns correspond with the observations made in survey analyses, which consistently find that $100, 50, 10, 5, 1, 0.5, 0.1, \ldots$ exhibit decreasing levels of roundedness (Battistin et al., 2003; Whynes et al., 2005; Covey and Smith, 2006). For these reasons, let me define the (degree of) focality of a number x as the level of the highest number in this sequence that divides x. The focality level of multiples of 100 is 4, other multiples of 50 have level 3, and so on; as defined in Table 3. Clearly, this notion of focality levels merely reflects the current understanding and leaves room for improvement in future work.[10]

Specification Adopting standard practice, all parameters $(\alpha, \beta, \lambda, \kappa)$ are heterogeneous across subjects in the sense that they differ randomly between subjects: α is truncated nor-

[10]Note that the distinction of discrete focality levels does not violate convexity of the image $\phi[X]$ assumed above, Assumption 1. The discrete focality function used here is defined for all multiples of 0.1 and can be extended straightforwardly to a continuous function on $\mathbb{R}$ obeying convexity of the image.

mal on $[-0.5, 0.5]$, β is normal, λ and κ are log-normal.[11] This mixed-logit model and the distributional assumptions are standard, see e.g. Bellemare et al. (2008), Wilcox (2008), and Train (2003).[12] Appendix A provides all details. Likelihoods are computed by numerical integration using quasi-random numbers (Train, 2003) and maximized in a three-step approach, using first a robust gradient-free approach (NEWUOA, Powell, 2006), secondly a Newton-Raphson method to ensure convergence, and finally extensive cross-testing of estimates across data sets and models to ensure that global maxima are found. Models are evaluated using the likelihood ratio test of Schennach and Wilhelm (2016) which is robust to both misspecification and arbitrary nesting of models. I will indicate significance of differences between models distinguishing the conventional level of 0.05 and the higher level of 0.005. The latter roughly implements the Bonferroni correction given the simultaneous tests of four models on four data sets, which I will focus on.

5 Results

First I analyze basic model adequacy, i.e. in-sample fit. The practically more relevant questions on counterfactual predictions, consistency of estimates and relevance are answered subsequently. Throughout, I relate the results for Focal to those of three key benchmark models. Besides the obvious benchmark "multinomial logit", which is the model used in most current analyses of discrete choice and hence must be considered, I consider a nested logit model allowing for similarity effects between proximate options (Ordered GEV, Small, 1987), to test for violation of IIA, and a model of limited attention following Echenique et al. (2014), which provides an alternative explanation for the focus on round numbers. The formal definitions of these models are provided in the appendix.

5.1 Capturing behavior

Table 4 summarizes the results on descriptive adequacy evaluated in terms of the Bayes information criterion (BIC) and the pseudo-R^2. The former quantifies model adequacy adjusting for the number of parameters used, the latter clarifies how much of the observed variance is explained in relation to the benchmarks *clairvoyance* and *naiveté*.[13] The relation signs indicate significance of differences, as defined in the table note. Table 4 pools the

[11]Further, the "payoffs" in the CES utility function are the experimental tokens earned by the subjects. The supplement provides extensive robustness checks for both assumptions. Allowing for subject heterogeneity in all dimensions allows for slightly more robust fit for all models, without obstructing identification. The token-based utility function used here fits best assuming the status quo model logit. The main alternative to the latter assumption is contextual utility (Wilcox, 2011), which tends to fit best for Focal and PALM, but these tendencies are overall minor and therefore not discussed here. The main results are fully robust to variations in these assumptions, as shown in the supplement.

[12]More generally, mixed logit models are standard in analyses of consumer demand, Berry et al. (1995), and in analyses of social preferences (Cappelen et al., 2007; Bellemare et al., 2008) as well as risk preferences (Harrison et al., 2007; Andersen et al., 2008).

[13]"Clairvoyance" predicts the choice distributions as they have been observed and "naiveté" predicts uniform randomization. Given a model's log-likelihood ll, $BIC = |ll| + \log(\#obs) \cdot \#par/2$ (Schwarz, 1978) and given the log-likelihoods of the "clairvoyant" model and the naive model, denoted as $ll_{\max}$ and $ll_{\min}$ respectively, the pseudo-R^2 is defined as $R^2 = (BIC - ll_{\min})/(ll_{\max} - ll_{\min})$ (Nagelkerke, 1991).

Table 4: Precision of the choice models in-sample (BIC: less is better; R^2: more is better)

	Value range		Focality		Similarity		Limited Attention		IIA
	Clairvoyance	Naiveté	(Focal)		(OGEV)		(PALM)		(Logit)
Large manual DG experiments (AM02, HJ06; 8 and 10 observations per subject)									
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 manual DG experiment (CHST07; 2 observations per subject)									
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 DG experiment (FKM07, 50 observations per subject)									
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

Content: For each choice model (Logit, OGEV, PALM, and Focal), the descriptive accuracy (BIC and pseudo-R^2 in-sample) is reported for data from the three groups of experiments. Significance of differences is evaluated by the robust Schennach and Wilhelm (2016) LR test: $\approx$ indicates p-values above 0.05, $>,<$ indicate p-values between 0.005 and 0.05, and $\gg,\ll$ indicate p-values below 0.005.

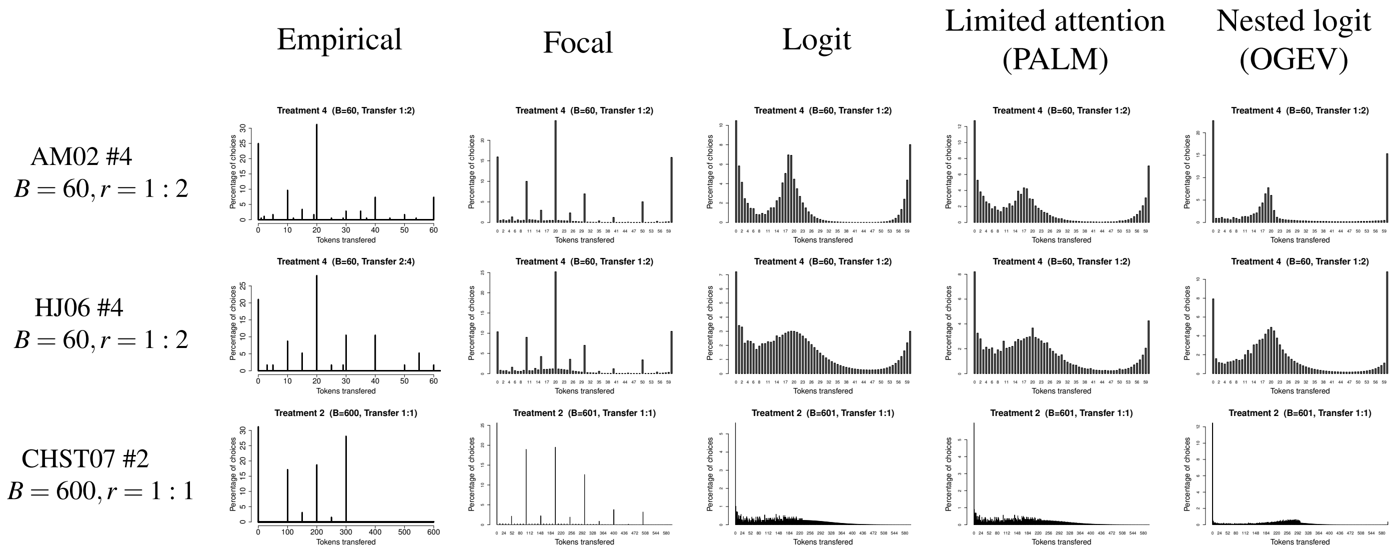
results for AM02 and HJ06 due to their similarity, both entailing choice from up to $B = 100$ options with 8 or 10 observations per subject. This way, Table 4 presents the results by type of data set: manual choice with many observations per subject (“large manual”), with few observations per subject (“small manual”), and graphical choice (“graphical”).

Table 4 shows that there are no substantial differences between models in the graphical DG but large differences in the manual DGs. This will be of relevance for utility estimation. Specifically, in the “large manual” experiments AM02 and HJ06, Focal captures 89% of observed variance, logit explains around 65% of variance, and PALM and OGEV improve slightly but statistically significantly on logit, explaining around 69% of observed variance. In the “small manual” experiment (CHST07), Focal’s adequacy is similar, at 83%, but the benchmark models all drop to around 35% in terms of the pseudo-R^2. In the graphical experiment, finally, all models explain around 62% of observed variance. The differences in the manual DGs are statistically highly significant, always at the 0.005 level surviving the Bonferroni correction.

Result 1 (Basic adequacy)**.** *Focal captures manual choice of numbers effectively, explaining 88% of the observed variance. All models capture graphical choice equally well.*

A drop of R^2 by 20 percentage points, from 89% to 69% as in AM02 and HJ06, or even from 83% to 35% as in CHST07, is substantial. In this sense, the benchmark models do not fit behavior “reasonably” well, and Figure 2 illustrates this by plotting the models’ predictions in reasonably representative treatments (a full list of plots is provided in the supplementary material). The main observation is that the benchmark models do not come close to capturing the highly discontinuous distribution of manually chosen numbers.

Figure 2: The precision of the choice models in capturing manual choice



Note: These plots depict the "predictions" of various models in the respective treatments after fitting the model parameters to all treatments from the respective experiments. The plots shown here represent a fairly representative selection, and the full list of plots is provided as supplementary material (as are the underlying parameter estimates). FKM07 is left out here, as Treatments as such are not defined (budget sets and transfer rates are individually randomized) and the differences between models are very minor in any case.

The reasons vary, but at least with hindsight they are clear[14], even the markedness of logit's inadequacy of capturing choice, which renders the popularity of logit all the more puzzling. Apparently, looking at data through the lens of logit (or similar models) yields a very "blurry" representation of behavior and can hardly yield useful utility estimates.

The in-sample results are intuitive also in other respects. First, Focal attains a substantially higher pseudo-R^2 in manual DGs than in graphical ones. This is intuitive, as the round-number pattern implies that choices cover relatively few options in manual DGs and thus indeed are more predictable than in graphical DGs. Second, the other models are equally adequate in the (large) experiments with and without round-number effects (AM02 and HJ06 vs. FKM07), confirming the optical impression that they do not comprehend round-number effects. Third, Focal's adequacy is largely similar in the large and small manual experiments, being 89% and 83%, respectively. These experiments differ in the relative strength of round-number effects: CHST07 implements the largest number of options, up to $B = 1600$, but the smallest number of different options actually gets chosen by the subjects (see also the discussion of entropy following shortly). In this sense, the round-number effects are strongest in CHST07, and the observation that Focal's ade-

[14]Utilities are continuous by assumption, which implies that logit's probabilities must be continuous. OGEV, while being devised as a general model of choosing from ordered option sets, primarily captures spikes at utility maximizers, most prominently zero transfers and Leontief transfers, which misrepresents that spikes actually relate to round numbers. Limited attention improves on logit in the "right" direction, being based on the focality index ϕ also used for Focal, but the effect is hardly visible. Round-number effects are much stronger than limited attention admits, and its limitation seems related to its axiom "Hazard Rate IIA" (Echenique et al., 2014), or "I-Asymmetry" and "I-Independence" in Manzini and Mariotti (2014).

Table 5: Is entropy reliably captured?

	Empirical	Focal		OGEV		PALM		Logit
All manual DGs	1.99	2.48*	$\ll$	3.58**	$\approx$	3.61**	$\approx$	3.63**
AM02 + HJ06	2.1	2.53	$<$	3.27**	$\approx$	3.29**	$\approx$	3.32**
CHST07	1.65	2.34	$<$	4.51**	$\approx$	4.57**	$\approx$	4.58**

Content: This table relates the empirical entropy (estimated entropy averaged across all treatments in the respective experiments) to the respective predictions of the four choice models. The results are either pooled across all manual DG experiments or reported for the large/small DG experiments (AM02 + HJ06 or CHST07, resp.). Differences are evaluated by Wilcoxon tests, the notation of relation signs is as in Table 4. The "stars" 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 *.

quacy is similar in both types of experiments indicates that it captures the round-number effect comprehensively. In turn, the adequacy of the other models drops substantially in CHST07, to around 35%, confirming that these models do not comprehend round-numbers effect: the stronger the round-number effect, the lower their adequacy. Finally, OGEV relaxes IIA and turns out to fit only slightly better than logit, which confirms the above hypothesis that similarity effects are of minor relevance in Dictator experiments.

Explaining entropy The most focal phenomenon related to round-number effects is the relatively low number of different options being chosen. This can be measured by estimating the entropy of the choice distribution. Using $\Pr(x)$ as the relative frequency of x, the Shannon-entropy $H = -\sum_{x:\Pr(x)\neq 0} \Pr(x)\log(\Pr(x))$ measures the information contained in a set of observations, and $\exp(H)$ quantifies the number of different options being chosen. To provide intuition, $\exp(H)$ is exactly 1 if all subjects choose the same option, it is equal to the number of options if the observations are distributed uniformly, and in the analyzed experiments, $\exp(H)$ is approximately equal to the number of different options that are (minimally) required to cover 90% of all choices. The estimates of $\exp(H)$ are around 5–10 in the manual DG experiments AM02, HJ06, CHST07, and around 30 in the graphical DG of FKM07.[15] Thus, subjects consistently focus on 5–10 options in manual DGs, although the total number of options ranges from 41 to 1601.

To evaluate whether choice is captured in this respect, I use the estimates obtained above and compute the predicted entropy in the various treatments. The significance of differences between predicted and observed entropies is evaluated in Wilcoxon matched pairs tests. The results are reported in Table 5 and confirm that both limited attention (PALM) and Ordered GEV slightly improve on logit in capturing the observed choice patterns, but are far from being compatible with the extent of the round-number effects. In turn, Focal is compatible with it in the sense that the predicted entropy, while being slightly too large,[16] is not significantly different from the observed entropy.

[15] Detailed overviews of these statistics are provided in the supplementary material. To compute these numbers for FKM07 (where budget sets are random between subjects), choices x_i are transformed to shares transfered, i.e. $x_i/\max x$, for each decision and rounded to multiples of 0.01.

[16] Intuitively, this relates to the noise in the data, which is unpredictable ex-ante but manifests as specific choices ex-post, and suggests that the assumed linear focality index may be improved upon.

Table 6: Mean degrees of altruism α and efficiency concerns β

	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

Content: This table reports the estimated means of the degree of altruism (α) and the degree of efficiency concerns (β). The estimates are given for each choice model (Logit, OGEV, PALM, Focal) and each data set. 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.

Result 2. *Focal captures the entropies of round-number choice across treatments.*

Finally, Table 6 provides the estimated mean degrees of altruism and efficiency concerns, i.e. the means of the distributions estimated for α and β across subjects. The mean degree of efficiency concerns (β) is similar across models and data sets, being close to zero (i.e. Cobb-Douglas). The mean degree of altruism varies substantially across data sets, however. The estimates of logit, PALM and OGEV range from roughly -0.1 to 0.5, which is volatile given the general bounds of -0.5 and 0.5. The extreme estimates are obtained in the experiments with the most pronounced round-number effects (i.e. with the lowest entropy), -0.1 in AM02 and 0.5 in CHST07. This shows that round-number effects do not simply bias estimates downwards or upwards. Their implications depend on the relation of round numbers to theoretically prominent choices such as Leontief. In contrast, Focal's estimates are fairly robust, ranging from 0.02 to 0.23. We next analyze whether these differences are significant and obstruct counterfactual predictions.[17]

5.2 Counterfactual predictions

Counterfactual predictions are a key component in applied work, as they are required for policy recommendations (on say tax rates) and to obtain ex-ante hypotheses for experiments. The following analysis will acknowledge that, when making predictions, we might have more or less information about the target environment. We might know "choice precision" of the group to be predicted (i.e. the distribution of λ), the "choice bias" associated with the interface used (distribution of κ), or nothing at all. To reflect these possibilities, and as robustness checks, I distinguish counterfactual reliability to three degrees. The first degree is the reliability of predicting preferences (mean and variance of α and β), while precision and choice bias (mean and variance of λ and κ) are known for the target environment. This evaluates the predictiveness of preference estimates in isolation. The second degree is the reliability of predicting both preferences and precision (mean and variance

[17]The full lists of parameter estimates, with robust standard errors, are provided in the supplementary material. Standard errors are skipped in Table 6, as they are not informative to evaluate significance of differences. It requires the joint evaluation of differences in four preference parameters and to control for both precision λ and choice bias κ. These concerns are addressed in likelihood ratio tests as described next, and as before, I use the robust LR test of Schennach and Wilhelm (2016).

Table 7: Counterfactual reliability of estimates (pseudo-R^2: more is better)

	Value range		Focality		Similarity		Limited Attention		IIA
	Clairvoyance	Naiveté	(Focal)		(OGEV)		(PALM)		(Logit)
First degree (Prediction of preferences)									
Large manual	1	0	0.846	$\gg$	0.522	$\gg$	0.486	$\gg$	0.443
Small manual	1	0	0.798	$\gg$	0.226	$\approx$	0.194	$\approx$	0.179
Graphical	1	0	0.587	$\gg$	0.456	$\approx$	0.413	$\approx$	0.425
Second degree (Prediction of preferences and precision)									
Large manual	1	0	0.85	$\gg$	0.511	$\gg$	0.486	$\gg$	0.443
Small manual	1	0	0.786	$\gg$	0.205	$\approx$	0.182	$\approx$	0.18
Graphical	1	0	0.58	$\gg$	0.43	$\approx$	0.4	$\approx$	0.411
Third degree (Prediction of preferences, precision and choice bias)									
Large manual	1	0	0.824	$\gg$	0.49	$\gg$	0.437	$\gg$	0.394
Small manual	1	0	0.757	$\gg$	0.205	$\approx$	0.172	$\approx$	0.184

Content: This table evaluates the accuracy of predicting behavior in either "large manual" experiments (D_{out} = AM02, HJ06), "small manual" experiment (D_{out} = CHST07) or "graphical" experiment (D_{out} = FKM07), using estimates from the respective other studies. The pseudo-R^2 are reported and relation signs indicate results of Schennach-Wilhelm likelihood ratio tests: $\approx$ indicates p-values above 0.05, $>, <$ indicate p-values between 0.005 and 0.05, and $\gg, \ll$ indicate p-values below 0.005.

of α, β, λ), while choice bias is assumed to be known, and the third degree requires out-of-sample prediction of all parameters (mean and variance of $\alpha, \beta, \lambda, \kappa$).

As counterfactuals, I use results from other experiments. That is, I take estimates from a given experiments, predict observations in any other experiment, and rotate such that each experiment is predicted based on estimates from any other experiment. Counterfactual reliability to the third degree entails prediction of choice bias κ, i.e. predicting the extent of round-number effects, which is not meaningful between manual and graphical experiments. There I focus on predictions between the manual experiments. In the following, I focus on reporting the pseudo-R^2, i.e. observed variance captured by the predictions, which is proportional to the BICs but numerically easier to interpret. The classes of experiments are labeled as above, large manual, small manual, and graphical.

Table 7 presents the results. The adequacy of the models for predictions is very similar to their adequacy in capturing behavior in-sample, but the differences between models increase. Predicting manual choice, Focal maintains 85% accuracy, while the accuracy of the benchmark models drops to 40%–50%. This shows that Focal's in-sample accuracy is not an artefact of overfitting—if anything, the other models had been overfit. The differences in predicting graphical choice are even more informative about the quality of parameter estimates. Graphical choice does not exhibit round-number effects and all models are equally adequate in-sample. Hence, all differences in predictions stem from inaccurate measurement of preferences in the original data sets. The substantially higher adequacy of Focal in predicting graphical choice, more than 10 percentage points out-of-sample where the difference was zero in-sample, clearly shows that Focal provides more accurate preference measurement and predictions.

Result 3. *Focal's counterfactual predictions are highly reliable, maintaining almost in-sample accuracy in both cases, graphical and manual choice entry, and robustly even if*

Table 8: Estimate consistency across experiments (differences in BIC in-sample and out-of-sample: less is better)

	Value range		Focality		Similarity		Limited Attention		IIA
	Clairvoyance	Naiveté	(Focal)		(OGEV)		(PALM)		(Logit)
First degree (Consistency of preferences)									
Large manual	0	22737.7	531.4	$\ll$	1601.6^{*}	$<$	2188.6^{**}	$>$	1655^{*}
Small manual	0	12910.4	1377.9^{*}	$\ll$	7111^{**}	$\ll$	8738.5^{**}	$\approx$	8890.6^{**}
Graphical	0	5606	137.4	$\ll$	859.6^{**}	$>$	733.7^{**}	$\approx$	730.4^{**}
Second degree (Consistency of preferences and precision)									
Large manual	0	22737.7	854.9	$\ll$	1825.5^{**}	$<$	2327.2^{**}	$>$	1755.7^{**}
Small manual	0	12910.4	1275.7^{*}	$\ll$	8108.3^{**}	$\ll$	9252.4^{**}	$\approx$	9359.2^{**}
Graphical	0	5606	151.4	$\ll$	908.3^{**}	$\gg$	660.4^{**}	$\approx$	686.7^{**}
Third degree (Consistency of preferences, precision and choice bias)									
Large manual	0	6447	235.5^{*}	$\ll$	806.3^{**}	$\approx$	746^{**}	$\approx$	703.3^{**}
Small manual	0	4765.1	667^{**}	$\ll$	1692.2^{**}	$\ll$	2262^{**}	$\approx$	2307.5^{**}

Content: This table evaluates consistency of estimates from either "large manual" experiments (D_{in} = AM02, HJ06), "small manual" experiment (D_{in} = CHST07) or "graphical" experiment (D_{in} = FKM07), in relation to estimates from the respective other studies. As before, $\approx$ indicates p-values above 0.05, $>,<$ indicate p-values between 0.005 and 0.05, and $\gg,\ll$ indicate p-values below 0.005. Further, "stars" indicate the significance of inconsistency; significance at 0.005 is indicated by ** and significance at 0.05 by *.

Note: The third degree does not involve predictions of the graphical experiment (FKM07). Thus, the numbers are not directly comparable to the other degrees.

we have to predict precision (λ) and choice bias (κ) of subjects in the target environment. Predicting manual choice, Focal's accuracy exceeds those of the benchmark models by at least 30 percentage points, and even on their in-sample fit, Focal's predictions improve by 15 percentage points. Predicting graphical choice, Focal improves on the benchmarks by at least 10 percentage points, despite the absence of round-number effects.

Finally, the reliability of counterfactual predictions is robust to limiting the knowledge about the target environment, i.e. when we move from degree 1 to degree 2 and degree 3. The precision of subjects (degree 2) is predicted reliably, i.e. it represents a robust facet of behavior, and Focal's accuracy drops only slightly (between two and four percentage points in the pseudo-R^2) if we have to predict the extent of the choice bias κ in addition (degree 3). In applications, neither precision nor choice bias therefore need to be known for the target environment. This shows that even the extent of round-number effects, as measured through Focal's κ, is a robust facet of behavior in manual choice experiments.

5.3 Consistency of estimates

Now assume we have two data sets and wish to examine if preferences in one experiment differ from those in the other one. To evaluate such "estimate consistency", the relevant likelihood ratio is the difference of log-likelihoods in-sample and out-of-sample. That is, we take estimates from a data set D_1, predict D_2, and compare the resulting likelihood to the one of the model when D_2 was estimated in-sample. Consistency is violated if the prediction fits significantly worse than the model had fit in-sample. Significance of consistency violation is tested again in the robust Schennach-Wilhelm test and indicated by asterisks in Table 8: one asterisk indicates weakly significant inconsistency (p-value below 0.05) and two asterisks indicate significance of inconsistency robust to the Bonferroni correction (p-value below 0.005). Similarly to above, I distinguish consistency to three degrees, consistency of preference estimates in isolation, consistency of preference and precision estimates jointly, and consistency of all estimates jointly, and also as above, I aggregate results based on user interface. Table 8 additionally reports on the relative consistency of the models, i.e. whether model estimates are significantly more or less consistent than those of other models (via relation signs), which I discuss further below.

The results in Table 8 confirm the above impression that Focal's predictions reach in-sample accuracy in most cases. Controlling for the different numbers of parameters in-sample and out-of-sample, the differences between in-sample and out-of-sample are mostly insignificant, violating consistency at the Bonferroni level of 0.005 in only $1/8$ cases. The estimates of logit, PALM and OGEV violate consistence at this level in $7/8$ cases each. This is informative, as consistency is in principle independent of the in-sample accuracy. For example, a model that does not fit well in-sample may still be particularly robust due to being "simpler" (Hey et al., 2010), e.g. by making less or fewer inadequate assumptions or by being less flexible and thus avoiding overfitting. Focal achieves both, capturing choice accurately in-sample as well as enabling reliable predictions out-of-sample and consistent estimates across samples—strongly suggesting that it provides an accurate description of choice across these standard experiments. This holds in particular in comparison with the benchmark models, as the relation signs in Table 8 show. A model is called significantly "more consistent" than another model if its inconsistency (difference out-of-sample and in-sample) is significantly lower than the inconsistencies of the other models, evaluated again in Schennach-Wilhelm LR tests. A single relation sign indicates significance at the 0.05 level, double relation signs indicate significance robust to the Bonferroni correction (0.005). Universally across the three groups of experiments and the three degrees of knowledge about the target environment, Focal's estimates are significantly more consistent than those of the other models at the robust 0.005 level.[18]

Result 4. *Preference estimates obtained via Focal from large experiments (at least eight observations per subject) do not significantly differ from those of other experiments, while those obtained from the small experiment (CHST07, two observations per subject) differ at least weakly significantly. The estimates obtained using the benchmark models generally differ highly significantly between all experiments.*

This shows that reliable preference measurement is possible, but it requires adequate models and a sufficient number of observations per subject. The former is an econometric concern, the latter is a concern about experimental design, and jointly they are to be observed to reliably detect behavioral differences across games and studies. Experimental designs that vary conditions mostly or only between subjects do not allow to reliably separate altruism (α), efficiency concerns (β) and precision (λ), implying that behavior is weakly identified in each dimension. Relatedly, let me clarify that the subject pools

[18] In turn, there is no robust ranking between logit, PALM and OGEV with respect to their consistency. Between these three models, every one of them is most consistent in one context and least consistent in another context.

are for our purposes identical: the counterfactual predictions (Table 7) show that estimates from other experiments allow to predict behavior in CHST07 accurately, close to achieving in-sample accuracy, but Table 8 shows that the estimates from CHST07 are not suitable to predict behavior in the other experiments. This clearly attests weak identification as opposed to differences between subject pools.

This observation corresponds also with the theoretical identification result. The key to identification was that the mapping from options to utilities or options to focalities varies across tasks, i.e. that utility and focality are not collinear, but that is not given in CHST07: the experiment always uses 1:1 transfers and only extends the option sets. All other experiments vary the transfer rates and thus change the mapping from options to utilities across tasks. Notably, Focal is still the least vulnerable model, but within-subject variation in choice tasks seems necessary to reliably disentangle preferences and noise, i.e. to detect behavioral differences across classes of games.

5.4 Comparative relevance

The relevance of controlling for heterogeneity in preference estimation is widely recognized. I use this observation to benchmark the relevance of controlling for presentation. Table 9 provides Bayes Information Criteria and pseudo-R^2 similar to above, but now as a function of the extent of subject heterogeneity being controlled for. The analysis focuses on logit and Focal, of which logit does not control for presentation effects while Focal by construction does so (the other benchmark models are very similar to logit, see above, and therefore skipped here). First, let us focus on the relevance in-sample, Table 9a. The "homogeneous logit" model neglects both subject heterogeneity and focality and captures 21.2% of observed variation. Controlling for heterogeneity allows to explain an additional 38% of observed variance, corroborating the intuition that controlling for heterogeneity is critical. Controlling for focality has the same impact in-sample, requiring only one additional parameter (κ) instead of four (variances of $\alpha, \beta, \lambda, \kappa$ to capture heterogeneity), and if applied in addition to controlling for heterogeneity, it still improves the R^2 by 30 percentage points. In-sample, that is, controlling for focality and controlling for heterogeneity are similarly important and almost perfectly complementary.

Tables 9b and 9c analyze reliability out-of-sample, and in particular, they analyze the reliability of the preference and precision estimates when making counterfactual predictions (degrees 1 and 2). The reliability of preference estimates (Table 9b) improves by around 49 percentage points in terms of R^2 by controlling for heterogeneity, and by another 43 percentage points on top controlling for focality. When predicting both preferences and precision in other experiments, the distributions of (α, β, λ) across subjects, controlling for heterogeneity improves upon the homogeneous logit model by 45 percentage points and controlling for focality adds another 44 percentage points on top of it (Table 9c).

Result 5. *Controlling for focality and controlling for heterogeneity are of similar relevance, highly complementary and both necessary in reliable preference measurement.*

The mixed logit model with full heterogeneity explains 59.8% percent of observed variance in sample, and it explains 40.1% out-of-sample; controlling for focality, the R^2 is

Table 9: Relevance of controlling for round-number effects in preference estimation, in relation to controlling for subject heterogeneity

(a) Accuracy in-sample (upper panel: BIC, lower panel: Pseudo-R^2; both, less is better)

	Clairvoyance	Naiveté	Repr Agent		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) Out-of-sample degree 1: Reliability of preference estimates (upper panel: BIC, lower panel: Pseudo-R^2)

	Clairvoyance	Naiveté	Repr Agent		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) Out-of-sample degree 2: Reliability of preference and precision estimates jointly (top: BIC, bottom: Pseudo-R^2)

	Clairvoyance	Naiveté	Repr Agent		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

Note: Each panel provides BIC and pseudo-R^2 for overall 4×2 models: four models of subject heterogeneity (starting with the "representative agent" model, i.e. homogeneity) and two models of choice (the status quo logit, and the generalization Focal). The underlying likelihood ratio tests follow Schennach and Wilhelm (2016), $>, <$ indicate p-values between 0.005 and 0.05, and $\gg, \ll$ indicate p-values below 0.005.

at a much higher level in-sample and drops substantially less out-of-sample, from 88.5% to 84%—showing that the reliability of estimates indeed improves. In contrast, the estimates of the logit model assuming homogeneity of subjects actually have negative external validity (-9.1% and -4.8% and Tables 9b and 9c), implying that representative-agent predictions provide at best uninformed guesses. This holds true whether or not we control for stochastic choice, and with hindsight, this is fairly obvious: The comparative statics depend on presentation, and if we calibrate a representative agent based on one set of observations, then the predictions for the other sets simply cannot be correct without controlling for representation. Since representative agent models are frequently used to theoretically predict behavior for policy recommendations and experimental hypotheses, this is a substantial concern and shows us that falsifying such predictions is not necessarily informative about the underlying model. The basic model used here, CES altruism, is fine, but the representative agent prediction neglecting presentation simply is inadequate.

6 Conclusion

This paper develops and applies a general framework allowing to capture presentation effects in stochastic choice. The point of departure is Gul and Pesendorfer's (2001) representation of set-based preferences, which captures the general notion that choice itself can be a non-trivial task and decision makers can be distracted (or "tempted") by suboptimal options. Their representation result is translated into invariance conditions about choice probabilities, augmented by the positivity axiom, and then solved to obtain a representation of stochastic choice allowing for distractions due to (e.g.) the focality of options. The model is tested by re-analyzing standard laboratory experiments on dictator games, and the results confirm the impression of the introductory examples. Choice patterns and comparative statics strongly depend on presentation, see Figure 1—and inadvertently, counterfactual predictions and utility analyses neglecting presentation are unreliable (Results 3 and 4). This implies that falsifying predictions or detecting behavioral differences between games without controlling for presentation risks being uninformative, if not misleading.

The idea that presentation of choice tasks affects choice seems widely recognized, as a number of reasons for presentation effects such as ordering, round-number, and left-digit effects are well-documented. Considering this, surprisingly few studies explicitly analyze presentation effects in choice, and the few examples I am aware of, e.g. Bernheim and Rangel (2007) and Salant and Rubinstein (2008), all focus on rational (non-stochastic) choice. This paper contributes to this literature by showing the high relevance, comparable to controlling for heterogeneity, and the possibility of rigorously modeling presentation effects allowing for stochastic choice, in both theoretical and experimental analyses.

The econometric results explicitly show that we can (and need to) factor in presentation when making predictions about experiments or policy interventions (Results 3 and 5) and that we can control for presentation effects when evaluating model adequacy across experiments (Result 4). In order to enable the latter, i.e. to identify models allowing for presentation and heterogeneity, we require within-subject variations of choice tasks (Result 4), namely to disentangle between-subject variation due to preference heterogeneity, within-subject variation due to noise, and within-subject variation due to presentation.

Without such designs, inference is unreliable and convergence of preference modeling is not attainable, simply due to the presentational differences between experiments. To give an idea about the magnitude of these differences, the designs used in analyses of dictator games range from experiments with two or three options to choose from (Charness and Rabin, 2002; Engelmann and Strobel, 2004) over experiments offering around 10 options (List, 2007; Dana et al., 2006) to experiments offering up to a hundred (Andreoni and Miller, 2002; Harrison and Johnson, 2006) or even a thousand of options (Fisman et al., 2007; Cappelen et al., 2007). Dictator experiments further range from allowing the equal split to result from choosing a round number to ruling out that the equal split is attainable at all, and from classroom experiments conducted with paper and pencil to computerized experiments allowing for graphical user interfaces or assistance in computing payoff profiles—and this is just the comparably small class of dictator games. Such differences are traditionally labeled "strategically irrelevant", but as shown above, they are choice relevant and substantially confound utility analyses. This confounds understanding behavior in dictator games and thus understanding differences to other games such as taking (List, 2007; Bardsley, 2008) and sorting games (Dana et al., 2006; Lazear et al., 2012), which in turn suggests that behavioral analyses may perpetually disagree unless we acknowledge presentation the way we acknowledge heterogeneity and stochastic choice.

Regarding implications for future work, in addition to allowing for within-subject variation of choice tasks, a few more points appear worth noting. First, the above analysis assumed that the relative focality of options is linear in the "level of roundedness" observed in statistical analysis of survey responses. This can be generalized straightforwardly, but linearity seems sufficient when analyzing dictator choice. Secondly, the existing experimental literature recognizes potential ordering effects in choice from lists and routinely reverses the ordering to nullify such effects. As discussed in Rubinstein and Salant (2006), the extreme options in lists may well have relatively high or relatively low focality, implying that order reversals are insufficient. An alternative approach implied by Theorem 2 is to arrange the options along a circle, which can be rotated, and the resulting equation system (see Section 3.3) identifies utility and focality up to affine transformation (Lemma 1 in the appendix). Finally, in two of the experiments considered here, the payoff-equalizing "Leontief" choices had been round numbers in all tasks (AM02 and CHST07), and these two experiments happened to yield estimates of utility parameters deviating the most from the graphical choice benchmark FKM07. It thus seems advisable to vary the roundedness of predictions associated with particular models (as in HJ06). Alternatively, the graphical interface of FKM07 mitigates round-number effects, but this "number free" choice elicitation is not applicable in experiments on strategic choice and difficult to use outside the laboratory, implying that presentation cannot be circumvented this way in general.

To conclude, there obviously are many open questions regarding presentation effects and their analysis. Given the possibility to study these effects theoretically and econometrically, as shown here, it seems possible to capture a wide range of such "choice distractions" in a rigorous and reliable manner. This is necessary yet also promising to advance our understanding of preferences across games, and hence warrants consideration and investigation in future work.

References

Aczél, J. and Dhombres, J. G. (1989). *Functional equations in several variables*. Cambridge university press.

Andersen, S., Harrison, G., Lau, M., and Rutström, E. (2008). Eliciting risk and time preferences. *Econometrica*, 76(3):583–618.

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.

Bardsley, N. (2008). Dictator game giving: Altruism or artefact? *Experimental Economics*, 11(2):122–133.

Battistin, E., Miniaci, R., and Weber, G. (2003). What do we learn from recall consumption data? *Journal of Human Resources*, 38(2):354–385.

Bellemare, C., Kröger, S., and van Soest, A. (2008). Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica*, 76(4):815–839.

Bellemare, C., Sebald, A., and Strobel, M. (2011). Measuring the willingness to pay to avoid guilt: estimation using equilibrium and stated belief models. *Journal of Applied Econometrics*, 26(3):437–453.

Benkert, J.-M. and Netzer, N. (2015). Informational requirements of nudging. Technical report, CESifo Group Munich.

Bernheim, B. D. and Rangel, A. (2007). Toward choice-theoretic foundations for behavioral welfare economics. *American Economic Review*, 97(2):464–470.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Birnbaum, M. H. and Navarrete, J. B. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty*, 17(1):49–79.

Breitmoser, Y. (2016). The axiomatic foundation of logit. *MPRA Paper 74334*.

Camerer, C. (1989). An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty*, 2(1):61–104.

Camerer, C., Ho, T., and Chong, J. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119(3):861–898.

Camerer, C. and Ho, T. H. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, 67(4):827–874.

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.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117(3):817–869.

Costa-Gomes, M., Crawford, V., and Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 69(5):1193–1235.

Covey, J. and Smith, R. (2006). How common is the âĂŸprominence effect'? Additional evidence to Whynes et al. *Health economics*, 15(2):205–210.

Dana, J., Cain, D., and Dawes, R. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, 100(2):193–201.

Davis, D. and Holt, C. (1993). *Experimental Economics*. Princeton University Press.

De Bruyn, A. and Bolton, G. E. (2008). Estimating the influence of fairness on bargaining behavior. *Management Science*, 54(10):1774–1791.

Dean, M. L. (1980). Presentation order effects in product taste tests. *The Journal of psychology*, 105(1):107–110.

Debreu, G. (1960). Review of 'Individual Choice Behavior' by R. Luce. *American Economic Review*, 50:186–8.

Dinner, I., Johnson, E. J., Goldstein, D. G., and Liu, K. (2011). Partitioning default effects: why people choose not to choose. *Journal of Experimental Psychology: Applied*, 17(4):332.

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

Engel, C. (2011). Dictator games: a meta study. *Experimental Economics*, 14:583–610.

Engelmann, D. and Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review*, 94(4):857–869.

Feenberg, D., Ganguli, I., Gaule, P., and Gruber, J. (2017). ItâĂŹs good to be first: Order bias in reading and citing nber working papers. *Review of Economics and Statistics*, 99(1):32–39.

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

Fudenberg, D., Iijima, R., and Strzalecki, T. (2015). Stochastic choice and revealed perturbed utility. *Econometrica*, 83(6):2371–2409.

Gächter, S., Herrmann, B., and Thöni, C. (2004). Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence. *Journal of Economic Behavior and Organization*, 55(4):505–531.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education LTD.

Gul, F., Natenzon, P., and Pesendorfer, W. (2014). Random choice as behavioral optimization. *Econometrica*, 82(5):1873–1912.

Gul, F. and Pesendorfer, W. (2001). Temptation and self-control. *Econometrica*, 69(6):1403–1435.

Harless, D. and Camerer, C. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, pages 1251–1289.

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.

Harrison, G. W., List, J. A., and Towe, C. (2007). Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion. *Econometrica*, 75(2):433–458.

Heitjan, D. F. and Rubin, D. B. (1991). Ignorability and coarse data. *Annals of Statistics*, pages 2244–2253.

Hey, J. (1995). Experimental investigations of errors in decision making under risk. *European Economic Review*, 39(3):633–640.

Hey, J. (2005). Why we should not be silent about noise. *Experimental Economics*, 8(4):325–345.

Hey, J., Lotito, G., and Maffioletti, A. (2010). The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity. *Journal of risk and uncertainty*, 41(2):81–111.

Keane, M. (2010a). A structural perspective on the experimentalist school. *Journal of Economic Perspectives*, 24(2):47–58.

Keane, M. and Wolpin, K. (2007). Exploring the usefulness of a nonrandom holdout sample for model validation: Welfare effects on female behavior. *International Economic Review*, 48(4):1351–1378.

Keane, M. P. (2010b). Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics*, 156(1):3–20.

Lacetera, N., Pope, D. G., and Sydnor, J. R. (2012). Heuristic thinking and limited attention in the car market. *American Economic Review*, 102(5):2206–2236.

Lazear, E. P., Malmendier, U., and Weber, R. A. (2012). Sorting in experiments with application to social preferences. *American Economic Journal: Applied Economics*, 4(1):136–163.

List, J. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, 115(3):482–493.

Luce, R. (1959). *Individual choice behavior: A theoretical analysis*. Wiley New York.

Manski, C. F. and Molinari, F. (2010). Rounding probabilistic expectations in surveys. *Journal of Business & Economic Statistics*, 28(2):219–231.

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.

Matejka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–98.

Mattsson, L.-G. and Weibull, J. W. (2002). Probabilistic choice and procedurally bounded rationality. *Games and Economic Behavior*, 41(1):61–78.

McFadden, D. (1974). Conditional logit analysis of qualitative choice models. *Frontiers of Econometrics, ed. P. Zarembka. New York: Academic Press*, pages 105–142.

McFadden, D. (1976). Quantal choice analysis: A survey. *Annals of Economic and Social Measurement*, 5(4):363–390.

McKenzie, C. R., Liersch, M. J., and Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science*, 17(5):414–420.

Miller, J. M. and Krosnick, J. A. (1998). The impact of candidate name order on election outcomes. *Public Opinion Quarterly*, 62:291–330.

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

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy*, 9(4):513–548.

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.

Poltrock, S. E. and Schwartz, D. R. (1984). Comparative judgments of multidigit numbers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1):32.

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

Read, D., Loewenstein, G., and Rabin, M. (1999). Choice bracketing. *Journal of Risk and Uncertainty*, 19(1-3):171–97.

Rubinstein, A. and Salant, Y. (2006). A model of choice from lists. *Theoretical Economics*, 1(1):3–17.

Rust, J. (2014). The limits of inference with theory: A review of wolpin (2013). *Journal of Economic Literature*, 52(3):820–850.

Salant, Y. and Rubinstein, A. (2008). (a, f): choice with frames. *The Review of Economic Studies*, 75(4):1287–1296.

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.

Starmer, C. and Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? an experimental investigation. *American Economic Review*, 81(4):971–978.

Thurstone, L. (1927). A law of comparative judgment. *Psychological review*, 34(4):273–286.

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

Wasserstein, R. L. and Lazar, N. A. (2016). The asa's statement on p-values: context, process, and purpose. *The American Statistician*.

Whynes, D., Philips, Z., and Frew, E. (2005). Think of a number… any number? *Health Economics*, 14(11):1191–1195.

Wilcox, N. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. In Cox, J. C. and Harrison, G. W., editors, *Risk aversion in experiments*, volume 12 of *Research in experimental economics*, pages 197–292. Emerald Group Publishing Limited.

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

A For online publication: Proofs and technical definitions

A.1 Proof of Theorem 1

Any choice profile Pr with a Focal representation satisfies Axioms 1–5, as can be verified easily. I focus on establishing $\Rightarrow$, showing that Axioms 1–5 imply that the choice profile has a Focal representation for some $\lambda, \kappa \in \mathbb{R}$. As in the paper, I write $u_x = u(x)$ and $\phi_x = \phi(x)$ for all $x \in X$. I say that u_x is interior if there exist $x', x'' \in X$ such that $u_{x'} < u_x < u_{x''}$.

Step 1: We show that for any given $(u, \phi) \in \mathcal{C}$, there exists $V_{u,\phi}\big(u_x, \phi_x\big)$ such that

$$\Pr(x|u, \phi, B) = \frac{V_{u,\phi}(u_x, \phi_x)}{\sum_{x' \in B} V_{u,\phi}(u_{x'}, \phi_{x'})} \qquad \text{for all } x \in B \in P(X).$$

Fix $(u, \phi) \in \mathcal{C}$. First, by positivity and IIA, for any $(u, \phi) \in \mathcal{C}$,

$$\Pr(x|u, \phi, B) = \frac{V\big(u_x, u_y, \phi_x, \phi_y, x, y\big)}{\sum_{x' \in B} V\big(u_{x'}, u_y, \phi_{x'}, \phi_y, x', y\big)} \qquad \text{for all } x, y \in B \in P(X), \tag{7}$$

with $V\big(u_x, u_y, \phi_x, \phi_y, x, y\big) := \Pr(x|u, \phi, \{x, y\}) / \Pr(y|u, \phi, \{x, y\})$. The argument, using solely IIA, is well-known, see e.g. McFadden (1974, p. 109). Eq. (7) holds true for all $x, y \in B$ and all $B \in P(X)$, i.e. for all $y \in X$. Hence, the odds of choosing x over x' are constant for any pair of benchmark options $y, y' \in X$,

$$\frac{\Pr(x|u, \phi, B)}{\Pr(x'|u, \phi, B)} = \frac{V(u_x, u_y, \phi_x, \phi_y, x, y)}{V(u_{x'}, u_y, \phi_{x'}, \phi_y, x', y)} = \frac{V(u_x, u_{y'}, \phi_x, \phi_{y'}, x, y')}{V(u_{x'}, u_{y'}, \phi_{x'}, \phi_{y'}, x', y')}$$

for all $x, x', y, y' \in B$ and all $B \in P(X)$. This ratio is therefore independent of y, and by convexity of X in $\mathbb{R}$ (richness) this implies that the derivate of the ratio with respect to any interior $y \in X$ is 0 and thus well-defined,

$$\frac{d}{dy} \frac{V(u_x, u_y, \phi_x, \phi_y, x, y)}{V(u_{x'}, u_y, \phi_{x'}, \phi_y, x', y)} = 0.$$

As a result, functions $f(y, u_y, \phi_y)$ and $V_1(x, u_x, \phi_x)$ exist such that $V(u_x, u_y, \phi_x, \phi_y, x, y) = V_1(x, u_x, \phi_x) \cdot f(y, u_y, \phi_y)$ for all $x, y \in X$, and we can write, for all $B \in P(X)$, $x \in B$ and $y \in X$,

$$\Pr(x|u, \phi, B) = \frac{V(u_x, u_y, \phi_x, \phi_y, x, y)}{\sum_{x' \in B} V(u_{x'}, u_y, \phi_{x'}, \phi_y, x', y)} = \frac{V_1(x, u_x, \phi_x)}{\sum_{x' \in B} V_1(x', u_{x'}, \phi_{x'})}.$$

Finally, "completeness" implies that for any x, x' such that $u_x = u_{x'}$ and $\phi_x = \phi_{x'}$, $V_1(x, u_x, \phi_x) = V_1(x', u_{x'}, \phi_{x'})$, which implies that the argument x can be dropped in the sense that there exists a function $V_{u,\phi} : \mathbb{R}^2 \to \mathbb{R}$ such that $V_{u,\phi}(u_x, \phi_x) = V_1(x, u_x, \phi_x)$ for all $x \in X$.

Step 2: $V_{u,\phi}(u_x,\phi_x) = \exp\{\lambda\, u_x + \kappa\, \phi_x + c_x\}$ for all $(u,\phi) \in \mathcal{C}$ and all $x \in X$.
By positivity, for any $(u,\phi) \in \mathcal{C}$, either $V_{u,\phi}(u_x,\phi_x) > 0$ for all x or $V_{u,\phi}(u_x,\phi_x) < 0$ for all x. Without loss of generality, assume $V_{u,\phi}(u_x,\phi_x) > 0$ for all x (otherwise, multiply $V_{u,\phi}$ by -1). By narrow bracketing and relative focality,

$$\frac{V_{u+r_u,\phi+r_\phi}(u_x+r_u,\phi_x+r_\phi)}{V_{u+r_u,\phi+r_\phi}(u_{x'}+r_u,\phi_{x'}+r_\phi)} = \frac{\Pr(x|u+r_u,\phi+r_\phi,B)}{\Pr(x'|u+r_u,\phi+r_\phi,B)} = \frac{\Pr(x|u,\phi,B)}{\Pr(x'|u,\phi,B)} = \frac{V_{u,\phi}(u_x,\phi_x)}{V_{u,\phi}(u_{x'},\phi_{x'})} \quad (8)$$

for all $(u,\phi) \in \mathcal{C}$, all $(r_u,r_\phi) \in \mathbb{R}^2$ and all $x,x' \in X$. Further, by "completeness", there exists a function $h : \mathbb{R}^2 \to \mathbb{R}$ such that, for all $(u,\phi) \in \mathcal{C}$, and all $(r_u,r_\phi) \in \mathbb{R}^2$,

$$\frac{V_{u+r_u,\phi+r_\phi}(u_x+r_u,\phi_x+r_\phi)}{V_{u+r_u,\phi+r_\phi}(u_{x'}+r_u,\phi_{x'}+r_\phi)} = \frac{h(u_x+r_u,\phi_x+r_\phi)}{h(u_{x'}+r_u,\phi_{x'}+r_\phi)} \qquad \text{for all } x,x' \in X.$$

Combined, by positivity of V, there exists a function $g : \mathbb{R}^2 \to \mathbb{R}$ such that

$$h(u_x+r_u,\phi_x+r_\phi) = V_{u,\phi}(u_x,\phi_x)\cdot g(r_u,r_\phi) = h(u_x,\phi_x)\cdot g(r_u,r_\phi),$$

for all $(u,\phi) \in \mathcal{C}$, $(r_u,r_\phi) \in \mathbb{R}^2$, and $x \in X$. By richness of choice tasks, for any $(p_1,p_2) \in \mathbb{R}^2$, there exist $(u,\phi) \in \mathcal{C}$ and $x \in X$ such that $p_1 = u_x$ and $p_2 = \phi_x$. Hence, the previous functional equation holds for all points in $\mathbb{R}^2$,

$$h(p_1+r_u,p_2+r_\phi) = h(p_1,p_2)\cdot g(r_u,r_\phi) \qquad \text{for all } (p_1,p_2),(r_u,r_\phi) \in \mathbb{R}^2,$$

By positivity, $h,g > 0$ implies that $\tilde{h} = \log h$ and $\tilde{g} = \log g$ are well-defined, which yields

$$\tilde{h}(p_1+r_u,p_2+r_\phi) = \tilde{h}(p_1,p_2) + \tilde{g}(r_u,r_\phi) \qquad \text{for all } (p_1,p_2),(r_u,r_\phi) \in \mathbb{R}^2. \qquad (9)$$

This is an instance of the fundamental Pexider functional equation in $\mathbb{R}^2$. Next, holding $(u,\phi) \in \mathcal{C}$ fixed and fixing an arbitrary benchmark option $y \in X$, from

$$\Pr(x|u,\phi,\{x,y\}) = \frac{\Pr(x|u,\phi,\{x,y\})}{\sum_{x' \in \{x,y\}} \Pr(x'|u,\phi,\{x,y\})} = \frac{V_{u,\phi}(u_x,\phi_x)}{\sum_{x' \in \{x,y\}} V_{u,\phi}(u_{x'},\phi_{x'})}$$

for all $x \in X$, it follows that $s \in \mathbb{R}$ exists such that $\Pr(x|u,\phi,\{x,y\}) = V_{u,\phi}(u_x,\phi_x)\cdot s$ for all $x \in X$. By positivity, $s > 0$. Now, $\Pr(x|u,\phi,\{x,y\}) \leq 1$ for all x implies $V_{u,\phi}(u_x,\phi_x) \leq 1/s$ for all $x \in X$ and therefore $h(u_x,\phi_x) \leq \log 1/s$ for all $x \in X$. Since the images $u[X],\phi[X]$ are convex, non-singleton subsets of $\mathbb{R}$, this implies that h is bounded from above on intervals of positive length (in each dimension, u_x and ϕ_x). Thus Theorem 9 (page 43) and Theorem 8 (page 17) in Aczél and Dhombres (1989) jointly imply that all solutions of the functional equation (9) satisfy

$$\tilde{h}(u_x,\phi_x) = \lambda u_x + \kappa \phi_x + c_x \qquad \text{for all } x \in X, (u,\phi) \in \mathcal{C},$$

for some $\lambda,\kappa \in \mathbb{R}$ and $c : X \to \mathbb{R}$. The parameters $\lambda,\kappa \in \mathbb{R}$ and $c : X \to \mathbb{R}$ are unique, as the images of u and ϕ are not collinear. Since Eq. (8) holds for all x,x', the weights λ and κ are independent of x, while c may be an arbitrary function $c : X \to \mathbb{R}$; also by Eq. (8), all these terms are independent of (u,ϕ). Thus, we obtain, using $V_{u,\phi}=\exp\tilde{h}$,

$$V_{u,\phi}(u_x,\phi_x)=\exp\{\lambda u_x+\kappa\phi_x+c_x\} \qquad \text{for all } x\in X,(u,\phi)\in\mathcal{C}. \tag{10}$$

Step 3: $V_{u,\phi}(u_x,\phi_x)=\exp\{\lambda\, u_x+\kappa\,\phi_x\}$.
For contradiction, assume the opposite, i.e. Eq. (10) and Axiom 5 are satisfied, but choice probabilities are not Focal, i.e. $c_x\neq$ const. Fix $x,y\in X$ such that $c_x\neq c_y$, $B=\{x,y\}$, a context $(u,\phi)\in\mathcal{C}$, and construct the "inverted" context (u',ϕ') such that

$$u_x=u'_y, \qquad u_y=u'_x, \qquad \phi_x=\phi'_y, \qquad \phi_y=\phi'_x.$$

By Assumption 1, an appropriate $(u',\phi')\in\mathcal{C}$ exists. By Axiom 5, we obtain

$$\frac{\Pr(x|u,\phi,B)}{\Pr(y|u,\phi,B)}=\frac{\Pr(y|u',\phi',B)}{\Pr(x|u',\phi',B)}, \tag{11}$$

which implies

$$\frac{\exp\{\lambda\cdot u_x+\kappa\cdot\phi_x+c_x\}}{\exp\{\lambda\cdot u_y+\kappa\cdot\phi_y+c_y\}}=\frac{\exp\{\lambda\cdot u'_y+\kappa\cdot\phi'_y+c_y\}}{\exp\{\lambda\cdot u'_x+\kappa\cdot\phi'_x+c_x\}}=\frac{\exp\{\lambda\cdot u_x+\kappa\cdot\phi_x+c_y\}}{\exp\{\lambda\cdot u_y+\kappa\cdot\phi_y+c_x\}}$$
$$\Leftrightarrow \qquad c_x-c_y=c_y-c_x \tag{12}$$

which implies $c_x=c_y$, the contradiction. Hence, c_x is const in x, implying that c cancels out, establishing $\Rightarrow$. □

A.2 Proof of Theorem 2

Fix $(u,\phi)\in\mathcal{C}$. If Pr has a Focal representation, we know from $v_x:=\log\Pr(x|X)$ that, for all $x\in B\in P(X)$,

$$\Pr(x|u,\phi,B)=\frac{\exp\{v_x\}}{\sum_{x'\in B}\exp\{v_{x'}\}}=\frac{\exp\{\lambda u_x+\kappa\phi_x\}}{\sum_{x'\in B}\exp\{\lambda u_{x'}+\kappa\phi_{x'}\}}.$$

Now define $a:X\to\mathbb{R}$ as $a(x)=v_x-\lambda u_x-\kappa\phi_x$ for all $x\in X$. Hence,

$$\Pr(x|u,\phi,B)=\frac{\exp\{v_x\}}{\sum_{x'\in B}\exp\{v_{x'}\}}=\frac{\exp\{\lambda u_x+\kappa\phi_x+a(x)\}}{\sum_{x'\in B}\exp\{\lambda u_{x'}+\kappa\phi_{x'}+a(x')\}}.$$

for all $x\in B\in P(X)$, and by transitivity, we obtain

$$\frac{\exp\{\lambda u_x+\kappa\phi_x\}}{\sum_{x'\in B}\exp\{\lambda u_{x'}+\kappa\phi_{x'}\}}=\frac{\exp\{\lambda u_x+\kappa\phi_x+a(x)\}}{\sum_{x'\in B}\exp\{\lambda u_{x'}+\kappa\phi_{x'}+a(x')\}}$$
$$\Rightarrow \quad \frac{\exp\{\lambda u_x+\kappa\phi_x\}}{\exp\{\lambda u_{x'}+\kappa\phi_{x'}\}}=\frac{\exp\{\lambda u_x+\kappa\phi_x+a(x)\}}{\exp\{\lambda u_{x'}+\kappa\phi_{x'}+a(x')\}} \quad \text{for all } x,x'\in X,$$

implying $a(x)/a(x') = 1$ for all $x, x' \in X$. Thus, there exists $a \in \mathbb{R}$ such that $v_x = \lambda u_x + \kappa\phi_x + a$ for all x. Now, to characterize a, define a sequence (x_ε) such that $\lim_{\varepsilon\to 0} v_{x_\varepsilon} = \inf_x v_x$, which implies $\lim_{\varepsilon\to 0}(\lambda u_{x_\varepsilon} + \kappa\phi_{x_\varepsilon}) = \inf_x(\lambda u_x + \kappa\phi_x)$ as a is constant. Since

$$\Pr(x|u,\phi,B) = \frac{\exp\{v_x - \inf v\}}{\sum_{x'\in B}\exp\{v_{x'} - \inf v\}} = \frac{\exp\{\lambda u_x + \kappa\phi_x - r\}}{\sum_{x'\in B}\exp\{\lambda u_{x'} + \kappa\phi_{x'} - r)\}},$$

for any $r \in \mathbb{R}$, we obtain for $a_1 := \inf_x v_x$ and $a_2 := \inf_x(\lambda u_x + \kappa\phi_x)$, by positivity

$$\lim_{\varepsilon\to 0}\frac{\Pr(x_\varepsilon|u,\phi,B)}{\Pr(x|u,\phi,B)} = \frac{\exp\{0\}}{\exp\{v_x - a_1\}} = \frac{\exp\{\lambda\cdot 0\}}{\exp\{\lambda u_x + \kappa\phi_x - a_2\}}$$

for all $x \in X$. Hence, $v_x - a_1 = \lambda u_x + \kappa\phi_x - a_2$, for all $x \in X$ and $(u,\phi) \in \mathcal{C}$, implying $a = a_1 - a_2 \equiv \inf_x v_x - \inf_x(\lambda u_x + \kappa\phi_x)$ as claimed. Lemma 1 shows that this implies identification of u and ϕ up to affine transformation from two adequate tasks. $\square$

A.3 Additional results clarifying identification

The results in this subsection are the basis of the discussion of identification in Section 3.3. Lemma 1 shows that both u and ϕ are identified (up to affine transformation) if we rotate options, implying that they indeed can be seen as distinct choice factors, and Lemma 2 shows that identification is not possible if we simply reverse the order of options.

Lemma 1. *Fix $B \in P(X)$ and any context (u,ϕ). Define a second context by rotating u toward $\tilde{u}$ such that*

$$\tilde{u}(x) = u(x-1) \text{ for all } x > \min B \qquad \text{and} \qquad \tilde{u}(\max B) = u(\min B).$$

Observations from the two contexts (u,ϕ) and $(\tilde{u},\phi)$ suffice to identify (u,ϕ) up to affine transformation.

Proof. Without loss of generality, assume $B = \{1,2,\ldots,10\}$. By Theorem 2, based on the choice probabilities observing in (u,ϕ) and $(\tilde{u},\phi)$, we obtain the equation system

$$\begin{aligned}
v_1 &= a + \lambda u_1 + \kappa\phi_1 && (1a) & \tilde{v}_1 &= \tilde{a} + \lambda u_{10} + \kappa\phi_1 && (1b)\\
v_2 &= a + \lambda u_2 + \kappa\phi_2 && (2a) & \tilde{v}_2 &= \tilde{a} + \lambda u_1 + \kappa\phi_2 && (2b)\\
&\vdots & & & &\vdots & &\\
v_{10} &= a + \lambda u_{10} + \kappa\phi_{10} && (10a) & \tilde{v}_{10} &= \tilde{a} + \lambda u_9 + \kappa\phi_{10}, && (10b)
\end{aligned}$$

where (v_i) and $(\tilde{v}_i)$ are known and $a, \tilde{a}, \lambda, \kappa, (u_i), (\phi_i)$ are unknown. The claim is that both (u_i) and (ϕ_i) are defined up to affine transformation.

First, rearrange the equation system by defining equations $(1c) = (1b) - (1a)$, $(2c) = (2b) - (2a)$ and so on, as well as equations $(1d) = (1a) - (2b)$, $(2d) = (2a) - (3b)$, and so on. This yields the following system.

$$v_1 - \tilde{v}_1 = a - \tilde{a} + \lambda u_1 - \lambda u_{10} \quad (1c) \qquad v_1 - \tilde{v}_2 = a - \tilde{a} + \lambda\phi_1 - \lambda\phi_2 \quad (1d)$$
$$v_2 - \tilde{v}_2 = a - \tilde{a} + \lambda u_2 - \lambda u_1 \quad (2c) \qquad v_2 - \tilde{v}_3 = a - \tilde{a} + \lambda\phi_2 - \lambda\phi_3 \quad (2d)$$
$$\vdots \qquad\qquad \vdots$$
$$v_{10} - \tilde{v}_{10} = a - \tilde{a} + \lambda u_{10} - \lambda u_9 \quad (10c) \qquad v_{10} - \tilde{v}_1 = a - \tilde{a} + \lambda\phi_{10} - \lambda\phi_1 \quad (10d)$$

Second, set $\phi_1 = 0$ and define equations $(1e) = (1d)$, $(2e) = (1d) + (2d)$, $(3e) = (2e) + (3d), \ldots, (10e) = (9e) + (10d)$, as well as constants $d_1 = v_1 - \tilde{v}_2$, $d_2 = v_2 - \tilde{v}_3$ and so on. This yields

$$\begin{aligned} d_1 &= a - \tilde{a} - \lambda\phi_2 && (1e)\\ d_1 + d_2 &= 2(a - \tilde{a}) - \lambda\phi_3 && (2e)\\ d_1 + d_2 + d_3 &= 3(a - \tilde{a}) - \lambda\phi_4 && (3e)\\ &\vdots && (13)\\ d_1 + \cdots + d_9 &= 9(a - \tilde{a}) - \lambda\phi_{10} && (9e)\\ d_1 + \cdots + d_{10} &= 10(a - \tilde{a}) - \lambda\phi_1 && (10e) \end{aligned}$$

Given $\phi_1 = 0$, equation (10e) defines $(a - \tilde{a})$, and using this, $\lambda\phi_2, \ldots, \lambda\phi_{10}$ are defined from equations (1e)–(9e).

Finally, set $u_1 = 0$, which implies that λu_{10} and λu_2 are defined from equations (1c) and (2c), since we know $a - \tilde{a}$. Knowing λu_2, equation (3c) identifies λu_3, which implies that (4c) identifies λu_4 and so on, up to λu_9 which is identified from (9c). □

Lemma 2. *Reversion of the order of options does not suffice for identification. That is, given the equation system*

$$v_1 = a + \lambda u_1 + \kappa\phi_1 \quad (1a) \qquad \tilde{v}_1 = \tilde{a} + \lambda u_{10} + \kappa\phi_1 \quad (1b)$$
$$v_2 = a + \lambda u_2 + \kappa\phi_2 \quad (2a) \qquad \tilde{v}_2 = \tilde{a} + \lambda u_9 + \kappa\phi_2 \quad (2b)$$
$$\vdots \qquad\qquad \vdots$$
$$v_{10} = a + \lambda u_{10} + \kappa\phi_{10} \quad (10a) \qquad \tilde{v}_{10} = \tilde{a} + \lambda u_1 + \kappa\phi_{10}, \quad (10b)$$

where (v_i) and $(\tilde{v}_i)$ are known and $a, \tilde{a}, \lambda, \kappa, (u_i), (\phi_i)$ are unknown, (u_i) and (ϕ_i) are not defined up to affine transformation.

Proof. For purpose of contradiction, assume (u_i) and (ϕ_i) are defined up to affine transformation. Without loss of generality, assume $u_1 = 0$ and $\phi_1 = 0$. Since u and ϕ are identified up to affine transformation, this implies that $\lambda u_j, \kappa\phi_j$ are uniquely identified by the above equation system for all $j = 2, \ldots, 9$. Now consider the equations

$$\begin{aligned} v_2 &= a + \lambda u_2 + \kappa\phi_2 && (2a)\\ v_9 &= a + \lambda u_9 + \kappa\phi_9 && (9a)\\ \tilde{v}_2 &= \tilde{a} + \lambda u_9 + \kappa\phi_2 && (2b)\\ \tilde{v}_9 &= \tilde{a} + \lambda u_2 + \kappa\phi_9 && (9b) \end{aligned}$$

and assume that a and $\tilde{a}$ are known. Then, this represents a system of four equations and four unknowns, $(u_2, u_9, \phi_2, \phi_9)$, and these unknowns are not contained in any other equation. Hence, they must be uniquely defined by these four equations. Note that these equations are either linearly dependent, if $(9b) = (2a) + (9a) - (2b)$, or inconsistent, if $(9b) \neq (2a) + (9a) - (2b)$, the contradiction. □

A.4 Modeling utility and subject heterogeneity

Subjects' utilities exhibit constant elasticity of substitution (CES) in the incomes (π_i, π_j),

$$u_i(\pi_i, \pi_j) = \left((1-\alpha)\cdot(1+\pi_i)^{\beta} + \alpha\cdot(1+\pi_j)^{\beta}\right)^{1/\beta}, \tag{14}$$

where α represents the degree of altruism and β the degree of efficiency concerns. The standard assumption is that incomes π_i, π_j represent incomes measured in experimental tokens. Alternatively, incomes may represent the monetary values underlying the experimental tokens or utilities may be normalized to the range $[0,1]$ as in $(u-u_{\min})/(u_{\max}-u_{\min})$. The latter has been proposed by Wilcox (2011), Padoa-Schioppa (2009), and Padoa-Schioppa and Rustichini (2014) based on neuro-economic evidence that stimuli adapt to the environment. Thus, the normalization may improve fit across treatments and across experiments. The supplementary material analyzes which of these approaches captures utility best. The differences in-sample are insignificant, and out-of-sample, token-based utilities tend to fit best for logit and OGEV, while contextual utility tends to fit best for PALM and Focal. The differences are fairly minor overall, but for clarity, the supplementary material reports robustness checks on all results for all three approaches to utility definition. The paper reports the results for the approach favoring the status quo (logit), i.e. token-based utility.

Each subject is characterized by precision λ, altruism α, efficiency concerns β, and choice bias κ. Subjects may be heterogeneous, i.e. all parameters may 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 $\mathbf{d}$ describes the choices o_s of subject $s \in S$ is

$$l(\mathbf{d}|o_s) = \int_{\mathbf{P}} f(\mathbf{p}|\mathbf{d})\cdot \Pr(o_s|\mathbf{p})\, d\,\mathbf{p}, \tag{15}$$

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 being analyzed. The integral in Eq. (15) is evaluated by simulation, using quasi random numbers following standard practice (Train, 2003). The underlying distributional assumptions are as follows: altruism α is normal truncated to the interval $[-0.5, 0.5]$, efficiency β is normal without truncation, precision λ and choice κ are log-normal. In each case, both mean and variance are considered free parameters of the model. Overall the models thus have (up to) eight free parameters, which is conservative in relation to regression models used in experimental analyses and in relation to the more progressive structural models (Harrison et al., 2007; Bellemare et al., 2008). Regardless, identifiability of the parameters is verified explicitly by analyzing reliability and consistency across experiments.

The supplementary material reports robustness checks investigating whether lower-dimensional models are possibly as adequate in-sample but more reliably identified and thus more robust out-of-sample. Heterogeneity of preferences is highly significant, as known from the literature and as indicated in Table 9. Heterogeneity of precision is similarly significant, both in-sample and out-of-sample. Heterogeneity of the choice bias κ is not significant in-sample but it is significant out-of-sample, i.e. allowing for heterogeneity of κ improves predictions for all "rich data sets" with at least eight observations per subject (AM02, HJ06, FKM07). In turn, identification is weak given the data from CHST07, as shown in the supplement, i.e. the two observations per subject do not allow us to disentangle preferences, precision and choice allowing for heterogeneity. For this reason, I skip predictions based on CHST07 in the main analysis, which allows me to use the most adequate model for the rich data sets allowing subjects to be heterogeneous in all four dimensions $(\alpha, \beta, \lambda, \kappa)$. Aggregating over subjects, the log-likelihood of the model is

$$ll(\mathbf{d}|o) = \sum_{s \in S} \log l(\mathbf{d}|o_s) \tag{16}$$

with $o = \{o_s\}_{s \in S}$. 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 (as described in the supplementary material).

Models are discriminated by likelihood ratio tests that allow for both misspecified and potentially nested models, following Schennach and Wilhelm (2016) and as described in the supplement. 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 most cases, due to multiple testing problems resulting from testing several models on several data sets simultaneously. By analyzing multiple data sets in parallel, I avoid the reliance on a single p-value, addressing the concerns of Wasserstein and Lazar (2016) and providing the general picture, but to account for the multiple testing problem, significance at the stricter level is generally focused on.

A.5 Benchmark models: Formal definitions

Logit DM with utility u chooses x with probability

$$\text{Logit:} \quad \Pr(x|u, \phi, B) = \frac{\exp\{\lambda \cdot u_x\}}{\sum_{x' \in B} \exp\{\lambda \cdot u_{x'}\}}. \tag{17}$$

Limited attention Round-number effects can be interpreted two ways: subjects either focus on some options or neglect other options. Masatlioglu et al. (2012) generalize revealed preference to account for DMs not considering all their options, Manzini and Mariotti (2014) generalize this idea to stochastic choice, and Echenique et al. (2014) generalize the model further by allowing for a weak "perception ordering": first all options at the highest perception level are considered, second the options at the next-highest level, and so on. This Perception Adjusted Luce Model (PALM) straightforwardly applies to focality effects, first the most focal options are considered, next the second layer, and so on, and hence it constitutes a natural benchmark for Focal. Formally, DM with utility u, focality ϕ, precision λ, and choice bias $\kappa \in [0,1]$, chooses $x \in B$ with

$$\text{PALM:} \qquad \Pr(x|u,\phi,B) = \mu(x,X) \cdot \prod_{k>\phi_x} \left(1 - \kappa \cdot \sum_{x' \in X : \phi_{x'} = k} \mu(x',X)\right) \tag{18}$$

where $\mu(x,X) = \text{Logit}(x) = \exp\{\lambda\, u_x\} / \sum_{x' \in X} \exp\{\lambda\, u_{x'}\}$. The focality index ϕ used here will (of course) be equivalent in Focal. While Echenique et al. define and axiomatize PALM only for $\kappa = 1$, I allow for the whole spectrum down to $\kappa = 0$ (which is logit). Further, I rescale the choice probabilities so they add up to 1, following Manzini and Mariotti's suggestion for cases without "outside options".

Similarity/Nested logit Choice violates IIA in the presence of "similarity" effects, and intuitively proximate numbers are more similar than distant numbers. Such similarity effects can be expressed by nested logit (McFadden, 1976) where DM first chooses a "nest" of options and secondly makes his final choice from this nest. Small (1987) introduces a cross-nested logit model (with overlapping nests) for choice from ordered sets, called *Ordered GEV*,[19] which intuitively captures possible similarity effects in manual choice. Here, DM first makes a tentative choice $y \in B$ and then reconsiders the neighborhood of y to make the final choice $x \in [y-w, y+w]$. To clarify the relevance of nesting and similarity effects, I include Ordered GEV as benchmark model. Formally, DM with utility u, precision λ, degree of correlation κ, bandwidth parameter $M < |X|$, and options represented by their integer ranks $s = 1,2,\ldots$, the choice probabilities are

$$\text{OGEV:} \quad \Pr(s) = \sum_{r=s}^{s+M} \frac{w_{r-s} \exp\left\{\lambda u_s / \kappa\right\}}{\exp\{I_r\}} \cdot \frac{\exp\{\kappa I_r\}}{\sum_{t=0}^{B+M} \exp\{\kappa I_t\}}$$

$$\text{with} \quad I_r = \ln \sum_{s' \in B_r} w_{r-s'} \exp\left\{\lambda u_{s'} / \kappa\right\}. \tag{19}$$

[19] All cross-nested logit models are compatible with random utility if utility perturbations have a generalized extreme value distribution (GEV), hence the name Ordered GEV.