Quantitative Economics 12 (2021), 251–281 1759-7331/20210251

Controlling for presentation effects in choice

Yves Breitmoser
Department of Business Administration and Economics, Bielefeld University

Experimenters make theoretically irrelevant decisions concerning user interfaces and ordering or labeling of options. Reanalyzing dictator games, I first show that such decisions may drastically affect comparative statics and cause results to appear contradictory across experiments. This obstructs model testing, preference analyses, and policy predictions. I then propose a simple model of choice incorporating both presentation effects and stochastic errors, and test the model by reanalyzing the dictator game experiments. Controlling for presentation effects, preference estimates become consistent across experiments and predictive out-of-sample. This highlights both the necessity and the possibility to control for presentation in economic experiments.

Keywords. Presentation effects, utility estimation, counterfactual predictions, laboratory experiment.

JEL classification. C10, C90.

1. Introduction

Many economic experiments analyze individual behavior to understand preferences, as understanding preferences is required for applied theoretical analyses and quantitative policy recommendations. Preferences appear to be contradictory across studies, however, which mitigates the reliability of counterfactual predictions and policy recommendations. Such contradictory results appear to be particularly prevalent in studies of social preferences. As I show here, a major reason for inconsistency between experiments appears to be that "presentation" of options affects choice. Presentation matters due to default effects (McKenzie, Liersch, and Finkelstein (2006), Dinner, Johnson,Goldstein, and Liu (2011)), left-digit effects (Poltrock and Schwartz (1984), Lacetera, Pope, and Sydnor (2012)), round-number effects (Heitjan and Rubin (1991), Manski and Molinari (2010)), and positioning effects (Dean (1980), Miller and Krosnick (1998), Feenberg, Ganguli, Gaule, and Gruber (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 incompatibility of observations across experiments and may also be a major reason for the failure to reach a consensus on (social) preference theories.

Yves Breitmoser: yves.breitmoser@uni-bielefeld.de

An earlier draft of the paper was circulated under the title "Discrete Choice with Presentation Effects." I am grateful to an anonymous reviewer, Niels Boissonnet, 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.

Available at http://qeconomics.org. https://doi.org/10.3982/QE1050

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.

In actual experiments, this seemingly trivial decision, cutting the cake into either 20 or 25 slices, has drastic behavioral implications. As this effect seems to have escaped attention so far, Figures 1 and 2 illustrate this in some detail. Table 1 below reviews the experiments I am reanalyzing here. A detailed discussion follows, but briefly, I focus on dictator game experiments that are behaviorally equivalent in the sense that if behavior depends solely on outcome-based preferences, then the observed choice patterns should be statistically indistinguishable. In each dictator game, each of the B tokens to be allocated by the dictator is worth ρ_1 monetary units to the dictator and ρ_2 monetary units to the recipient. I refer to ρ_2/ρ_1 as the transfer ratio, where $\rho_2/\rho_1 > 1$ indicates subsidized transfers and $\rho_2/\rho_1 < 1$ indicates taxed transfers. Such dictator games are widely used to study social preferences, but specific experimental designs differ in theoretically irrelevant ways.

First, Andreoni and Miller (2002, AM02) allow for budgets up to $B = 100$ where the equal split is generally a multiple of 10 or 25, and choices mainly are multiples of 10 (Figure 1(a)). Similarly, Cappelen, Hole, Sørensen, and Tungodden (2007, CHST07) allow for budgets up to $B = 1600$ with round-numbered equal splits, and subjects primarily choose multiples of 100 (Figure 1(b)). Second, Harrison and Johnson (2006, HJ06) allow for budgets up to $B = 100$, but the equal split is often a plain integer or not attainable at all; as a result, it is chosen less often and choices overall are more dispersed (Figure 1(c)). Third, Fisman, Kariv, and Markovits (2007, FKM07) implemented a graphical user interface preventing round-number effects (Figure 1(d)). There, choices vary continuously and the equal split is not chosen more frequently than other options. The choice of either of these three types of experimental designs thus affects the relative frequency of the equal split, the dispersion of choices—and as Figure 2 shows, even the comparative statics. Figure 2 displays the share of tokens transferred as a function of the transfer ratio (tax or subsidy rate) for the three experiments actually varying ρ_1 and ρ_2. In HJ06, transfers are generally high and independent of ρ_2/ρ_1, in AM02 transfers start low but are increasing in ρ_2/ρ_1, and in FKM07 transfers are low and again independent of ρ_2/ρ_1.

To summarize, even in theoretically equivalent experiments, choice patterns differ widely and elicited comparative statics end up being contradictory—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

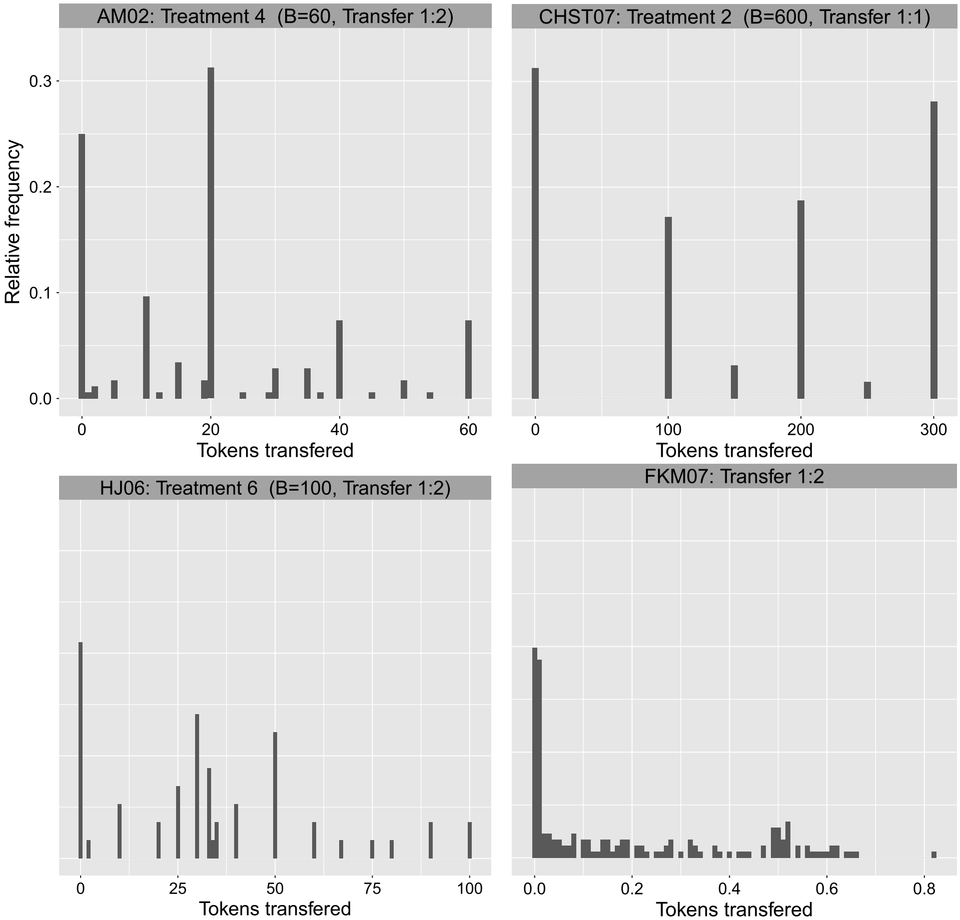

FIGURE 1. Choice patterns across dictator game experiments. *Note*: In AM02, the equal split is a round number (transferring 20 tokens), in CHST07 as well (300 tokens), in HJ06 it is not an integer and thus unavailable (transferring 33.3 tokens), and FKM07 use a graphical user interface. As above, B indicates the number of tokens available, and "transfer" (e.g., "1 : 2") indicates the inverse transfer ratio $\rho_1 : \rho_2$ as defined in the text.

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 proposes a simple 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 control 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 evi-

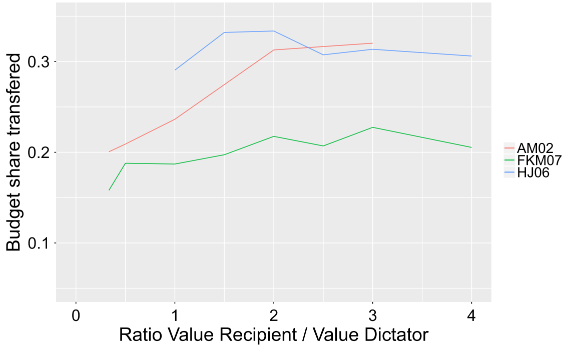

FIGURE 2. Comparative statics of transfer in tax/subsidy rate across experiments. *Note*: The plot captures the budget share transferred as a function of the transfer ratio in AM02, FKM07, and HJ06. The transfer ratio ρ_2/ρ_1 ranges from 1/4 (highly taxed transfers) to 4/1 (highly subsidized transfers).

dence 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, and by doing so, we can get much further in analyzing experimental studies than we previously thought.

Section 2 introduces the focal choice adjusted logit (Focal) model. Section 3 presents the framework for the econometric test of the model. Section 4 presents the results and Section 5 concludes. The Appendix contains relegated proofs and definitions, the Online Supplemental Material is available in the replication file (Breitmoser (2021)) and contains results on a number of robustness checks.

2. The model

The purpose of the model proposed here is to provide a framework for the estimation of utility parameters given an experimenter's hypothesis about the utility function. To obtain data, the experimenter runs an experiment where each decision maker DM chooses an option $x \in B$ from a finite budget set $B \subset X$. Each potential option $x \in X$ is associated with a list of attributes, and according to the experimenter's model, a subset of those attributes affect DM's utility. The remaining attributes do not affect DM's utility (theoretically) but may otherwise affect choice. In the existing literature, such utility-irrelevant yet choice-relevant attributes have been associated with temptation (Gul and Pesendorfer (2001)), salience (Bordalo, Gennaioli, and Shleifer (2012)) and focality (Schelling (1960)) of options, as discussed below.

Considering my focus on presentation effects, I will say that the utility-irrelevant attributes affect the *focality* of options, which refers to the "focal points" discussed by Schelling (1960) that reflect the "prominence or conspicuousness" (p. 57) of options. To be clear, Schelling discussed focal points in coordination games, where actions may gain prominence or conspicuousness from common pieces of knowledge or cultural background, but since presentation equally affects prominence or conspicuousness of options, it seems appropriate to borrow the term for this paper.

The set of possible budgets in the experiment is the set of finite subsets of X, denoted as $P(X)$. Further, $\pi : X \to \mathbb{R}^m$ denotes the mapping from options to values of theoretically utility-relevant attributes and $\phi : X \to \mathbb{R}^n$ denotes the mapping from options to values of other attributes such as ordering, positioning, labeling, and roundedness, which typically are considered utility irrelevant.[1] I denote the respective attribute values of option x as π_x and ϕ_x.

Example 1. DM chooses an option $x \in \{1, 2, 3\}$, the payoff of which depends on a coin toss (with unknown probabilities). Option 1 yields payoffs of 10 and 0 in the cases of heads and tails, respectively, option 2 yields 4 and 4, and option 3 yields 0 and 10. The options are presented in a column, with 1 being the top option (arbitrarily labeled "1") and 3 being at the bottom (arbitrarily labeled "0"). Thus:

$$\pi_1 = (10, 0), \qquad \phi_1 = 1 \mathrel{\hat{=}} \text{"Top"},$$
$$\pi_2 = (4, 4), \qquad \phi_2 = 0.5 \mathrel{\hat{=}} \text{"Middle"},$$
$$\pi_3 = (0, 10), \qquad \phi_3 = 0 \mathrel{\hat{=}} \text{"Bottom"}.$$

The images of π and ϕ are denoted $\pi[X]$ and $\phi[X]$, respectively. We assume that DM has complete, transitive, and reflexive orderings over both sets of attribute values, $\pi[X]$ and $\phi[X]$, intuitively interpreted as "preferred to" and "more focal than." The orderings are represented by a *utility index* $u : \pi[X] \to \mathbb{R}$ and a *focality index* $v : \phi[X] \to \mathbb{R}$, respectively.

Example 2. DM is risk neutral and believes the coin is fair, $u(\pi_x) = 0.5\pi_{x,1} + 0.5\pi_{x,2}$, and she has a tendency to choose options further down the list, $v(\phi(x)) = (1 - \phi(x))^2$.

Our task will be to infer utilities $u[\pi[X]]$ and focalities $v[\phi[X]]$ of options from DM's choice probabilities. In this task, the numeric labels attached to "Top," "Middle," and "Bottom" in the example are indeed arbitrary.

The pair of attribute mappings (π, ϕ) is called *context* of DM's choice and may vary between decision tasks, for example, by experimenters changing outcomes π or permuting the order ϕ. The set of possible contexts (π, ϕ) is denoted as $\mathcal{C}$. The probability that DM chooses $x \in B$ from budget $B \in P(X)$ in context $(\pi, \phi) \in \mathcal{C}$ is denoted as $\Pr(x|\pi, \phi, B)$.

As discussed in detail next, DM seeks to maximize $u + v$ subject to random perturbations of both utilities and focalities. This yields the following model.

Definition 1. The choice profile Pr has a focal choice adjusted logit (Focal) representation if there exist $\lambda, \kappa \in \mathbb{R}$, $u : \pi[X] \to \mathbb{R}$, and $v : \phi[X] \to \mathbb{R}$ such that for all contexts

[1]Presentation is obviously more complex than just comprising ordering, positioning, labeling, and roundedness of numbers. Take, for example, the way questions are asked in surveys. For my model to be applicable, we would have to be able to categorize (or quantify) the way questions are asked, so we can define a function mapping each category to focalities of different options. To the extent that such categories can be defined objectively, the model seems applicable, but I leave this for future research.

$(\pi, \phi) \in \mathcal{C}$ and all budgets $B \in P(X)$,

$$\Pr(x|\pi, \phi, B) = \frac{\exp\{\lambda \cdot u(\pi_x) + \kappa \cdot v(\phi_x)\}}{\sum_{x' \in B} \exp\{\lambda \cdot u(\pi_{x'}) + \kappa \cdot v(\phi_{x'})\}}. \tag{1}$$

If $\mathcal{C}$ is a singleton set, that is, comprises just a single context, then Pr has a Focal representation if and only if it has a logit representation. The indices u and v may therefore be identified only if we have observations from multiple contexts. Next, I briefly review related literature and then discuss identification. The Appendix contains a simple axiomatic characterization of the model.

Related literature The idea to distinguish utility-relevant and utility-irrelevant (but still choice-relevant) attributes of options when modeling choice has a long tradition in behavioral economics, presumably starting with Schelling (1960) who discussed focality of options as a choice relevant input in coordination games. Schelling's ideas have been picked up frequently again (for a discussion see Kreps (1990)), though with little attention to formalizing the notion of focality (but see Sugden (1995)). The idea of focal points, that is, prominence or conspicuousness of options, naturally extends to individual decision making, as evidenced by the well-documented effects due to ordering, labeling, or positioning of options or roundedness of numbers. The above approach, to model the interaction of utility and focality based on two indices u and v, that affect decision making in an additively separable manner ($u + v$) follows three strings of results in the literature on individual decision making.

First, additive separability of two such indices was proposed in Gul and Pesendorfer (2001) based on an axiomatic characterization of a preference for commitment. Formally, GP01 model a decision maker who, in a first stage, selects the menu of options to choose from, and in a second stage, an option from this menu. GP01's preference for commitment is a weak preference of DM in stage 1 for removing certain options from her eventual menu in stage 2. GP01 have phenomena typically related to time inconsistency in mind, for example, removing the option to have a burger at tomorrow's lunch, but the model itself is general and speaks to any phenomenon related to decision making that involves option attributes that are choice relevant but not utility relevant. To see this, note that if there exist option attributes that do not affect utility but are relevant to choice, then GP01's preference for commitment naturally arises. Such attributes may include attributes affecting Gul and Pesendorfer's notion of temptation, but critically for our discussion, they may also include attributes affecting (Schelling's) focality of options—formally, temptation as in GP01 and focality as in Schelling equally imply a preference for commitment. The key result of Gul and Pesendorfer is that, under seemingly standard assumptions, the preference for commitment implies that DM is influenced by two indices u and v, seeking to maximize $u + v$ when making the decision, as we assumed above—with the only difference that we call v "focality" index rather than "temptation" index.

Here, u represents the binary preference relation over singleton sets of options (hence, "utility" of options) and is revealed by the choice of a menu in the first stage.

Further, $u + v$ represents the preferences over options for any given menu of options and is revealed by the choice of an option in the second stage. Thus, anticipating that she will be tempted to choose options that do not maximize her utility u (due to the influence of v when making the decision), DM may prefer to eliminate some options from the menu prior to making her actual choice. In typical experimental designs, DM has no control over the menu of options and we get to observe DM only in Gul and Pesendorfer's second stage, where she seeks to maximize $u + v$. Not observing the menu preferences complicates identification, but in general, even knowing both u and $u + v$ (up to ordinal transformation) is not sufficient to infer v up to ordinal transformation (Gul and Pesendorfer (2006)).

As a second approach, Matejka and McKay (2015) modeled a DM with rational inattention and show that with attention costs following Sims (2003), choice probabilities obey a generalized logit form,

$$\Pr(x|\pi, \phi, B) = \frac{\exp\{\lambda u_x + \alpha_x\}}{\sum_{x' \in B} \exp\{\lambda u_{x'} + \alpha_{x'}\}}$$

for some $\lambda \in \mathbb{R}$, some utility index u and a "prior index" $\alpha : X \to \mathbb{R}$. That is, based on a conceptually independent analysis of stochastic choice, Matejka and McKay also obtain the additive separability earlier derived by GP01. In their model, α captures the choice propensities if DM has no knowledge of utility-relevant attributes, and thus, α again captures utility-irrelevant but choice-relevant attributes in choice, potentially related to temptation, salience, or focality of options, for example. Their results suggest that we may identify u and α (or, v) if we can present the decision problem without revealing utility relevant attributes, as the elicited choice probabilities then reveal α. This approach is feasible if the utility-relevant option attributes are known and have an empty intersection with the attributes affecting α, but may in general be hard to implement.

Third, in an analysis of the axiomatic foundation of conditional logit, Breitmoser (2020) observed that the same generalized logit representation, including the additive separability of utilities and "prior propensities," obtains if choice probabilities obey positivity, independence of irrelevant alternatives, and invariance of choice probabilities to translation of utilities. Based on this observation, an axiomatic foundation of the Focal model is straightforward, as demonstrated in the Appendix. As in Matejka and McKay, the prior propensities α are shown to be independent of utilities, that is, α remains unchanged if we permute or transform utilities in any way. Since utilities as such are not observable, such permutations are not straightforward, but in any model-based analysis, there are theoretically utility-relevant attributes and we may be able to permute them without affecting focality-relevant attributes. This would enable identification, as demonstrated next.

Identification As indicated, I envision an analyst seeking to estimate the parameters of a specific utility function controlling for specific presentation effects due to ordering, positioning or round numbers. I illustrate a prototypical example of such an application below by re-analyzing a set of dictator games that form the backbone of many studies of social preferences and exhibit the widely documented round-number effects. In these cases, the attributes affecting utility are specified prior to the analysis, but secondary attributes such as ordering of options (Example 1) need to be set to run the experiment, and their impact is supposed to be factored out. Then it will be possible to change the secondary attributes (say, by reordering options) without affecting the theoretically utility-relevant attributes, and equally, circular permutations of utility-relevant attributes are possible.

Definition 2. $\tilde{\pi}$ is a circular permutation of π with respect to $B \in P(X)$ if there exists a function $\sigma : B \to B$ such that: (i) $\tilde{\pi}_{\sigma(x)} = \pi_x$ for all $x \in B$, and (ii) $\sigma^k(x) \neq x$ for all $x \in B$ and all $k < |B|$.

As indicated, I propose to think of this assumption as clarifying that we are able to reposition, reorder or relabel options without affecting their utilities, which is true by all received models of utility in the behavioral literature. Obviously, such an assumption may become inadequate as our understanding of utility and presentation evolves, and in this sense, the present paper merely makes a first step in formally capturing presentation effects, but as we shall see, with respect to the experiments reanalyzed here, this assumption seems to be adequate. Circular permutations are sufficient for identification, as the following result demonstrates.

Proposition 1. *Fix a budget $B \in P(X)$ and consider two contexts $(\pi, \phi), (\tilde{\pi}, \phi) \in \mathcal{C}$, assuming DM's choice profiles have Focal representations with the same $\lambda, \kappa \in \mathbb{R}$ in both. Then utility index $u \circ \pi$ and focality index $v \circ \phi$ over options $x \in B$ are identified up to affine transformation if $\tilde{\pi}$ is a circular permutation of π with respect to B.*

That is, we can disentangle $u \circ \pi$ and $v \circ \phi$, that is, the mappings from options to utilities and from options to focalities, using appropriate experimental designs. Since π and ϕ are induced by the experimenter, this can be used for inference about functional forms and parameters of u and v. At this point, I am not aware of experiments actually implementing circular permutations, as opposed to say reversions of lists on which I comment below, and thus, I am not able to illustrate the approach based on existing data. The proposition is therefore a "possibility result," that is, a demonstration that it can be straightforward to disentangle u and v using experiments that vary u and v in a sufficiently informative way. Indeed, the experiments I reanalyze below do not seem to permit circular permutations in an obvious manner, as they require subjects to enter numbers rather than picking items from lists, but they generally provide observations from more than two contexts per subject, which at least for the experiments I am reanalyzing seems to provide sufficiently informative variation of π and ϕ (based on inspection of the gradients and Hessians of the likelihood function).

Briefly, let me illustrate how we can disentangle u and v using circular permutation of options. Assume that we do not know u and v from Example 1, but that we observe choice probabilities in the context (π, ϕ) defined in Example 1 and in the context $(\tilde{\pi}, \phi)$ defined next.

Example 3. Equivalent to Example 1, but utility-relevant attributes are given by $\tilde{\pi}$ as follows:

$$\tilde{\pi}_1 = (0, 10), \qquad \tilde{\pi}_2 = (10, 0), \qquad \tilde{\pi}_3 = (4, 4).$$

Note that $\tilde{\pi}$ in Example 3 is a circular permutation of π as defined in Example 1. By Proposition 1, observing choice probabilities in the two contexts, (π, ϕ) from Example 1 and $(\tilde{\pi}, \phi)$ from Example 3, suffices to identify the three utilities $u(\pi_1), u(\pi_2), u(\pi_3)$, and the three focalities $v(\phi_1), v(\phi_2), v(\phi_3)$ up to affine transformation. The basic idea is as follows. Knowing that the choice probabilities have a Focal representation with constant λ, κ in both contexts, we can demonstrate that the log-probabilities satisfy

$$\log \Pr(x|\pi, \phi, X) = a + \lambda u(\pi_x) + \kappa v(\phi_x) \tag{2}$$

for all $x \in \{1, 2, 3\} = B$ and both contexts, (π, ϕ) and $(\tilde{\pi}, \phi)$, with (λ, κ) as in the Focal representation and some $a \in \mathbb{R}$. In total, we have $2 \cdot |B| = 6$ such equations and $2 \cdot |B| + 4$ unknowns: the three utilities, the three focalities, λ and κ, and the two constants labeled a above (one per context). By fixing levels and scales of both utilities and focalities, we eliminate four unknowns, implying that the number of observations equates with the number of unknowns. The resulting equation system is regular if $\tilde{\pi}$ is a circular permutation of π. After fixing levels and scales of utilities and focalities, they are thus identified up to affine transformation.

Circular permutations are feasible in many cases. For example, many experimental analyses assume that utilities are functions of payoff profiles over states of the world or over players, and they allow subjects to choose one of n options. These options are arranged arbitrarily on a screen, or are associated arbitrarily with labels, and the analyst is worried that positioning or labeling of options biases choice. Any such arrangement permits a circular permutation, where say option 1 assumes the position (or, label) of option 2, option 2 that of 3, …, and option n assumes that of option 1. If the analyst observes choice probabilities for two such arrangements, where one is a circular permutation of the other, then identification up to affine transformation is possible.

There are two related points worth noting. On one hand, not all permutations of π yield a regular equation system. Most prominently, experimental studies involving lists of options tend to revert the list in order to nullify ordering effects by pooling observations from the original list and the reverse list. I am not aware of a possible model of choice justifying this approach, and as an example, assume both first and last option have high focality and all other options have low focality—then reversion has no effect on choice probabilities and neither utility nor focality is identified. Using the Focal model, one may define an equation system as above, but the resulting equation system is in general singular. In this case, identification of utilities using the Focal model can be achieved only if the ordering effect (i.e., v) is linear in the position of the option (ϕ), while circular permutation works in general.

On the other hand, we do not require assumptions about the functional form relating utility-relevant attributes and utilities of options in order to infer utilities of options, or about the functional form relating focality-relevant attributes and focalities in order

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, Kariv, and Markovits (2007))	50	500–1000	76×50	$4:1,\ldots,1:3$

Note: For each of the experiments reanalyzed here, the table lists the number of decisions per subject (#Treatments), the number of options per choice task (#Options), the number of observations overall (#Observations, which is the number of subjects times the number of treatments), and the range of transfer ratios implemented across treatments in the experiment.

to infer focalities of options. That is, circular permutation permits nonparametric estimation of cardinal utilities controlling for stochastic choice and presentation effects if we can observe choice probabilities. An analyst interested in estimating parameters may still do so, of course, but it is not necessary to estimate cardinal utilities.

3. Framework for model test and application

One approach to test the Focal model would be to test underlying axioms (see the Appendix) in isolation, that is, in specifically designed choice tasks. Such a test would be of limited informativeness here, as violations of axioms appear to be context dependent, suggesting that an applied model such as Focal is tested ideally based on data representative of those it is supposed to be applied to. I therefore test Focal on data sets representative of some of the behavioral literature, and attempt to answer three questions that are particularly relevant 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 economically substantial? The general idea, to test behavioral models by evaluating validity in-sample and out-of-sample using typical data sets, follows a literature comprising analyses of decision under risk (Harless and Camerer (1994), Wilcox (2008), Hey, Lotito, and Maffioletti (2010)), learning (Camerer and Ho (1999)), strategic choice in normal-form games (Camerer, Ho, and Chong (2004)), and social preferences (De Bruyn and Bolton (2008)).

The selected data sets are from the four well-known experiments on generalized dictator games discussed in the Introduction, reviewed in Table 1.

DEFINITION 3 (Generalized dictator game). DM chooses an option $x \in \{0, 1, \ldots, B\}$. Given x, the dictator's payoff is $\pi_1(x) = \rho_1 \cdot (B - x)$ and the recipient's payoff is $\pi_2 = \rho_2 \cdot x$.

Dictator games enable an analysis of utility and focality in a context where utility and focality individually appear to be understood reasonably well and additional factors can be ruled out thanks to the simplicity of dictator games.

On one hand, most analyses of dictator games that estimate utility parameters, including three of the four studies reanalyzed here, assume that preferences have a CES utility representation.[2] In this sense, assuming CES utilities fits my objective of reanalyzing standard data sets under standard assumptions, with the sole novelty being the attempt to control for presentation effects using the Focal model.

Definition 4 (Utility). The utility of option x, and given the associated payoff profile $(\pi_{x,1}, \pi_{x,2})$, is $u(\pi_{x,1}, \pi_{x,2}) = ((1-\alpha)\cdot(1+\pi_{x,1})^{\beta} + \alpha\cdot(1+\pi_{x,2})^{\beta})^{1/\beta}$.

On the other hand, there are unambiguous presentation effects (round-number effects), the choice task does not involve confounds due to, for example, risk, probability weighting, or (strategic) uncertainty, and there exist four well-known experimental analyses designed to estimate utility functions and differing only in presentational aspects. Further, there is consensus on the roundedness of numbers from analyses of surveys, which consistently find that $100, 50, 10, 5, 1, 0.5, 0.1, \ldots$ exhibit decreasing levels of roundedness (Battistin, Miniaci, and Weber (2003), Whynes, Philips, and Frew (2005), Covey and Smith (2006)). This eliminates another unknown and allows me to define the focality of option x as the level of the highest number in this sequence that divides x. The focality of multiples of 100 is 4, other multiples of 50 have level 3, and so on, down to multiples of 0.1, which have focality -2 and represent the minimum here.[3]

Definition 5 (Focality). Let ϕ_x denote the number that is to be entered in order to choose option x. Let $d : \mathbb{Z} \to \mathbb{R}$ be defined such that $d(0) = 1$, $d(1) = 5$ and $d(z) = 10\cdot d(z-2)$ for all $z \in \mathbb{Z}$. The focality of option x is $v(\phi_x) = \max\{z \mid d(z) \text{ divides } \phi_x\}$.

Finally, one of these experiments (Fisman, Kariv, and Markovits (2007)) avoids presentation effects due to round numbers by using a graphical user interface, which provides "counterfactual" information when analyzing consistency and reliability.[4] In conjunction, these four large-scale experiments therefore appear particularly suitable for testing Focal, as they provide a fairly comprehensive picture of choice in a context that is otherwise well understood.

[2] The exception is Cappelen et al. (2007), who allow for nontrivial reference points in order to analyze social norms. Relatedly, these studies also found that individual choices are frequently consistent with GARP, potentially suggesting that modeling stochastic choice is not needed in analyses of dictator games. All four of the studies reanalyzed here allow for stochastic choice, however, since the choices of the majority of subjects, sometimes the vast majority as in Fisman, Kariv, and Markovits (2007), are not perfectly consistent with any particular family of utility functions and since the standard approaches toward capturing population heterogeneity, such as finite mixture or mixed logit modeling, require a notion of stochastic choice to evaluate log-likelihoods in the first place. Note that the necessity for capturing population heterogeneity does not arise in studies analyzing behavior at the individual level (see, e.g., Halevy, Persitz, and Zrill (2018)).

[3] Note that negativity of focality has no particular meaning in the Focal model, as level shifts cancel out.

[4] Since the graphical user interface prevents round-number effects, on which I focus here, my analysis will assume that focality is constant at zero for all options in Fisman, Kariv, and Markovits (2007). Further work may investigate if the graphical user interface induces prominence effects other than via round numbers, at for example midpoints, which is possible by evaluating the significance of parameters of corresponding focality functions.

Specification Each subject is characterized by precision λ, altruism α, efficiency concerns β, and focality weight κ. Subjects may be heterogeneous, which we model by allowing for all parameters to be distributed randomly across subjects. Using $\mathbf{p} = (\lambda, \kappa, \alpha, \beta)$ to describe the parameter profile of a given subject and $f(\cdot|\mathbf{d})$ to describe its joint density in the population given distribution parameters $\mathbf{d}$, the likelihood that the model $\mathbf{d}$ describes the choices o_s of subject $s \in S$ is

$$l(\mathbf{d}|o_s) = \int_{\mathbf{P}} f(\mathbf{p}|\mathbf{d}) \cdot \Pr(o_s|\mathbf{p})\, d\mathbf{p}, \tag{3}$$

with $\Pr(o_s|\mathbf{p})$ as the probability that o_s results under parameter profile $\mathbf{p}$. Aggregating over subjects, the log-likelihood of the model given data set $o = \{o_s\}_{s\in S}$ is

$$ll(\mathbf{d}|o) = \sum_{s\in S} \log l(\mathbf{d}|o_s). \tag{4}$$

As usual, the underlying distributions are variations of normal distributions: altruism α is normal truncated to the interval $[-0.5, 0.5]$ to avoid excessive altruism or spite, efficiency β is normal without truncation, precision λ and choice κ are log-normal to avoid negativity. From a wider perspective, this specification closely relates to the widely-used family of mixed-logit models and the distributional assumptions reflect standard practice in analyses of social preferences (Cappelen et al. (2007), Bellemare, Kröger, and van Soest (2008)) and risk preferences (Harrison, List, and Towe (2007), Andersen, Harrison, Lau, and Rutström (2008), Wilcox (2008)).

For each of the random variables, 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, List, and Towe (2007), Bellemare, Kröger, and van Soest (2008)). Regardless, identifiability of the parameters is verified explicitly by analyzing reliability and consistency across experiments. Parameters are estimated by maximum likelihood. 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)), second, 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 always indicate the actual p-values, but let me note that a level of 0.005 roughly implements the Bonferroni correction given the simultaneous tests of four models on three data sets (in relation to an original level of 0.05).

Benchmark models Throughout the analysis, I relate the results for Focal to those for three key benchmark models for the lack of comparable earlier studies. An obvious benchmark is "multinomial logit," which is the model used in most current analyses.

Accordingly, DM with utility u chooses x with probability

$$\text{Logit}: \quad \Pr(x|u,\phi,B) = \frac{\exp\{\lambda \cdot u_x\}}{\sum_{x' \in B} \exp\{\lambda \cdot u_{x'}\}}. \tag{5}$$

In addition, I consider a cross-nested logit model allowing for *similarity* effects between proximate options (Ordered GEV, Small (1987)) and a model of *limited attention* following Echenique, Saito, and Tserenjigmid (2018), which provides an alternative explanation for the focus on round numbers.

Similarity 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 similar options and secondly makes a final choice from this nest. Small (1987) introduces a cross-nested logit model (with overlapping nests) for choice from ordered sets, called *Ordered GEV*, 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

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

$$\text{with } I_r = \ln \sum_{s' \in B_r} w_{r-s'} \exp\{\lambda u_{s'}/\kappa\}. \tag{6}$$

In the analysis, I use a large bandwidth w equal to half the available options, weights following a standard normal distribution, and free parameters λ to capture precision and κ to capture the extent of the choice bias due to similarity.

Limited attention Round-number effects can be interpreted two ways: subjects either focus on some options (as in Focal) or neglect other options, possibly due to limited attention. Masatlioglu, Nakajima, and Ozbay (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, Saito, and Tserenjigmid (2018) 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

$$\text{PALM}: \quad \Pr(x|u, \phi, B) = \mu(x, X) \cdot \prod_{k>\phi_x} \left(1 - \kappa \cdot \sum_{x' \in X : \phi_{x'} = k} \mu(x', X)\right), \tag{7}$$

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 to the one used in Focal. While Echenique, Saito, and Tserenjigmid 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."

4. Results

For brevity, I pool the 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, reducing the data to three types of experiments: manual choice with many observations per subject ("large manual", AM02 and HJ06), with few observations per subject ("small manual," CHST07), and graphical choice ("graphical," FKM07).

In-sample accuracy First, I analyze descriptive adequacy, that is, in-sample fit. It will be convenient to have the standard formal definition as a reference when discussing other measures of adequacy below. Using $\mathbf{d} = (d_\alpha, d_\beta, d_\lambda, d_\kappa)$ as the vector of distribution parameters with, for example, $d_\alpha = (\mu_\alpha, \sigma_\alpha)$ characterizing the distribution of α in the population, the descriptive log-likelihood with respect to data set o is

$$LL_o^{\text{descr}} = \max_{(d_\alpha, d_\beta, d_\lambda, d_\kappa)} ll\big((d_\alpha, d_\beta, d_\lambda, d_\kappa)|o\big),$$

for all data sets $o \in O := \{\text{AM02, HJ06, CHST07, FKM07}\}$, using the log-likelihood function defined in equation (4). Based on this, we define the descriptive Bayes information criterion (BIC, Schwarz (1978)) and the pseudo-R^2 (Nagelkerke (1991)) as usual,

$$BIC_o^{\text{descr}} = \left|LL_o^{\text{descr}}\right| + \log(\#\text{obs}) \cdot \#\text{par}/2, \qquad R_o^2 = \frac{BIC_o - LL_o^{\min}}{LL_o^{\max} - LL_o^{\min}}, \tag{8}$$

where $LL_o^{\max}$ denotes the log-likelihood of a (hypothetical) "clairvoyant" model predicting the choice distributions as they have been observed, and $LL_o^{\min}$ denotes the log-likelihood of a naive model predicting uniform randomization in all conditions. The first panel of Table 2 summarizes the results on in-sample accuracy. As indicated, I pool the (similar) "large manual" experiments AM02 and HJ06 in terms of their BIC, reporting the sum $BIC_{\text{AM02}}^{\text{descr}} + BIC_{\text{HJ06}}^{\text{descr}}$, based on which the pooled pseudo-R^2 is computed.

In these "large manual" experiments (AM02 and HJ06), Focal captures 89% of observed variance in terms of pseudo-R^2, logit captures 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% in terms of the pseudo-R^2, but the benchmark models drop to 35% in terms of the pseudo-R^2. In the graphical experiment, 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 (Model adequacy). *Focal captures manual choice of numbers in-sample, explaining* 88% *of the observed variance. All models capture graphical choice equally well.*

A drop of R^2 by 20 percentage points between Focal and the benchmark models, from 89% to 69% as in AM02 and HJ06, is substantial, and a drop from 83% to 35% as in CHST07 is drastic. Figure 3 below illustrates this by plotting the models' predictions in representative treatments, showing that the benchmark models do not capture choice "reasonably" well. Notably, the benchmark models are equally adequate in the (large) experiments with and without round-number effects (AM02 and HJ06 vs. FKM07), suggesting that they simply fail to capture the round-number clustering. Focal captures round-number clustering, and thus capitalizes on the fact that choice in manual DGs is more predictable. The remainder examines if this actually yields more accurate or reliable utility estimates.

Counterfactual predictions As for counterfactual predictions, there is a vast set of applications ranging from policy recommendations (where we may have little ex ante information) to hypotheses for experiments building on earlier experiments (where we may have much prior information about say precision and choice patterns). In order to account for this diversity, I distinguish predictions to three degrees: there is no information at all (*third degree*), there is information about the extent of the choice bias κ (*second degree*), or there is information about both precision and choice bias (*first degree*). In turn, the first degree thus evaluates the predictiveness of preference estimates in isolation, taking the distributions of λ and κ in the target population as known.

The counterfactuals to be predicted in my analysis are the choices in other experiments. I estimate parameters on a given experiment, predict observations in the other experiments, and rotate such that each experiment is predicted based on estimates from each other experiment. As outlined above, in predictions to the first degree ("Prediction of preferences" in Table 2), the utility parameters are out-of-sample and precision λ as well as choice bias κ are assumed to be known for the sample to be predicted—to measure the reliability of preference estimates in isolation. I implement this measure by taking the utility parameters from the other sample as given and estimating the distributions of λ and κ for the sample to be "predicted." Formally, when "predicting preferences" in data set (experiment) $o \in O := \{\text{AM02}, \text{HJ06}, \text{CHST07}, \text{FKM07}\}$, I take for each $o' \neq o$ the estimated distribution parameters $(d_\alpha^{o'}, d_\beta^{o'})$ of utilities, then maximize the log-likelihood over (d_λ, d_κ) with respect to data set o, and finally aggregate over all data sets $o' \neq o$,

$$LL_o^{\text{pred-1}} = \sum_{o' \neq o} LL_{o|o'}^{\text{pred-1}} \quad \text{with } LL_{o|o'}^{\text{pred-1}} = \max_{(d_\lambda, d_\kappa)} ll\big((d_\alpha^{o'}, d_\beta^{o'}, d_\lambda, d_\kappa)|o\big). \tag{9}$$

Table 2. Summary of the econometric analysis (BIC: less is better; R^2: more is better).

		Value Range								
		Clairvoyance	Naiveté	Focality (Focal)		Similarity (OGEV)		Limited Attention (PALM)		IIA (Logit)
Accuracy in-sample (BIC and corresponding pseudo-R^2 based on LL_o^{descr}, see equation (8))										
Large manual	BIC	2812.1	8137	3371.9	$<$ (0)	4420.9	$\approx$ (0.108)	4492.6	$<$ (0)	4690.3
	R^2	1	0	0.895	$>$ (0)	0.698	$\approx$ (0.108)	0.684	$>$ (0)	0.647
Small manual	BIC	260.8	1271.4	430.4	$<$ (0)	920	$\approx$ (0.733)	910.5	$\approx$ (0.567)	931.1
	R^2	1	0	0.832	$>$ (0)	0.348	$\approx$ (0.733)	0.357	$\approx$ (0.567)	0.337
Graphical	BIC	10021.8	23249.2	15103.9	$\approx$ (0.891)	15087.7	$\approx$ (0.754)	15119.4	$\approx$ (0.821)	15123.9
	R^2	1	0	0.616	$\approx$ (0.891)	0.617	$\approx$ (0.754)	0.615	$\approx$ (0.821)	0.614
Counterfactual predictions based on equations (9)–(11)										
Prediction of preferences (pseudo-R^2 based on $LL_o^{\text{pred-1}}$)										
Large manual	R^2	1	0	0.846	$>$ (0)	0.522	$>$ (0)	0.486	$>$ (0)	0.443
Small manual	R^2	1	0	0.798	$>$ (0)	0.226	$\approx$ (0.326)	0.194	$\approx$ (0.589)	0.179
Graphical	R^2	1	0	0.587	$>$ (0)	0.456	$\approx$ (0.063)	0.413	$\approx$ (0.694)	0.425
Prediction of preferences and precision (pseudo-R^2 based on $LL_o^{\text{pred-2}}$)										
Large manual	R^2	1	0	0.85	$>$ (0)	0.511	$>$ (0.005)	0.486	$>$ (0)	0.443
Small manual	R^2	1	0	0.786	$>$ (0)	0.205	$\approx$ (0.446)	0.182	$\approx$ (0.879)	0.18
Graphical	R^2	1	0	0.58	$>$ (0)	0.43	$\approx$ (0.193)	0.4	$\approx$ (0.706)	0.411
Prediction of preferences, precision and choice bias (pseudo-R^2 based on $LL_o^{\text{pred-3}}$)										
Large manual	R^2	1	0	0.824	$>$ (0)	0.49	$>$ (0)	0.437	$>$ (0)	0.394
Small manual	R^2	1	0	0.757	$>$ (0)	0.205	$\approx$ (0.339)	0.172	$\approx$ (0.827)	0.184
Consistency of estimates as defined in equations (12)–(14)										
Consistency of preference estimates										
Large manual	ΔBIC	0	22737.7	$531.4^{(0.272)}$	$<$ (0)	$1601.6^{(0.01)}$	$<$ (0.008)	$2188.6^{(0.001)}$	$>$ (0.016)	$1655^{(0.007)}$
Small manual	ΔBIC	0	12910.4	$1377.9^{(0.009)}$	$<$ (0)	$7111^{(0)}$	$<$ (0)	$8738.5^{(0)}$	$\approx$ (0.478)	$8890.6^{(0)}$
Graphical	ΔBIC	0	5606	$137.4^{(0.1)}$	$<$ (0)	$859.6^{(0)}$	$>$ (0.036)	$733.7^{(0)}$	$\approx$ (0.622)	$730.4^{(0)}$
Consistency of preferences and precision										
Large manual	ΔBIC	0	22737.7	$854.9^{(0.118)}$	$<$ (0)	$1825.5^{(0.004)}$	$<$ (0.04)	$2327.2^{(0.001)}$	$>$ (0.028)	$1755.7^{(0.005)}$
Small manual	ΔBIC	0	12910.4	$1275.7^{(0.014)}$	$<$ (0)	$8108.3^{(0)}$	$<$ (0)	$9252.4^{(0)}$	$\approx$ (0.596)	$9359.2^{(0)}$
Graphical	ΔBIC	0	5606	$151.4^{(0.073)}$	$<$ (0)	$908.3^{(0)}$	$>$ (0)	$660.4^{(0)}$	$\approx$ (0.157)	$686.7^{(0)}$
Consistency of preferences, precision and choice bias										
Large manual	ΔBIC	0	6447	$235.5^{(0.006)}$	$<$ (0)	$806.3^{(0)}$	$\approx$ (0.181)	$746^{(0)}$	$\approx$ (0.158)	$703.3^{(0)}$
Small manual	ΔBIC	0	4765.1	$667^{(0)}$	$<$ (0)	$1692.2^{(0)}$	$<$ (0)	$2262^{(0)}$	$\approx$ (0.281)	$2307.5^{(0)}$

Note: *Explanation*: The table provides results for "large manual" experiments (AM02 and HJ06 in aggregation), the "small manual" experiment (CHST07), and the "graphical" experiment is FKM07. It provides measures for both BIC and pseudo-R^2 (see text). We distinguish fit in-sample (both BIC and pseudo-R^2), out-of-sample (reporting pseudo-R^2), and consistency (difference of out-of-sample BIC and in-sample BIC). All details are provided in the text.
The six columns list the "clairvoyant" $LL_o^{\max}$ of a hypothetical model (with zero-free parameters) predicting the choice distributions as observed, the minimal $LL_o^{\min}$ of a hypothetical model predicting uniform randomization in each treatment, and the results for the four models defined in Sections 2 and 3. Relation signs indicate the direction of differences and the p-values evaluated by the robust Schennach and Wilhelm (2016) LR test are provided in parentheses underneath the relation signs. Regarding the consistency measure ΔBIC, the superscripts indicate the p-values of testing the null that the in-sample BIC equals the out-of-sample BIC (or equivalently, that ΔBIC is zero).

In predictions to the second degree ("Prediction of preferences and precision" in Table 2), also the distribution parameters of λ are taken out-of-sample,

$$LL_o^{\text{pred-2}} = \sum_{o' \neq o} LL_{o|o'}^{\text{pred-2}} \quad \text{with } LL_{o|o'}^{\text{pred-2}} = \max_{d_\kappa} ll\big(\big(d_\alpha^{o'}, d_\beta^{o'}, d_\lambda^{o'}, d_\kappa\big)|o\big), \tag{10}$$

and in predictions to the third degree ("Prediction of preferences, precision, and choice bias" in Table 2), predictions are fully out-of-sample,

$$LL_o^{\text{pred-3}} = \sum_{o' \notin \{o, \text{FKM07}\}} LL_{o|o'}^{\text{pred-3}} \quad \text{with } LL_{o|o'}^{\text{pred-3}} = ll\big(\big(d_\alpha^{o'}, d_\beta^{o'}, d_\lambda^{o'}, d_\kappa^{o'}\big)|o\big). \tag{11}$$

Counterfactual predictions to the third degree entail prediction of choice bias κ, that is, 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. Using these log-likelihoods, I compute BIC (e.g., $BIC_o^{\text{pred-1}}$ and $BIC_{o|o'}^{\text{pred-1}}$) and pseudo-$R^2$ as above, noting that the penalty term in BIC is a function of the number of free parameters actually fitted to the data set that is predicted (e.g., zero free parameters in predictions to the third degree).

The second panel of Table 2 reports the results. For brevity, I focus on the pseudo-R^2, that is, 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. Overall, the adequacy of the models for predictions is very similar to their adequacy in capturing behavior in-sample, but the differences between models increase further. Predicting manual choice, Focal maintains 85% accuracy, showing that Focal did not overfit in-sample, while the accuracy of the benchmark models drops to 40%–50%, suggesting they did overfit. Predicting graphical choice, where all models capture 62% of variance in-sample, Focal stays at 58%, while the benchmark models drop to 45%. These differences in predicting graphical choice are perhaps most informative about the quality of preference estimates, as 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.

Result 2. *Focal's counterfactual predictions are highly reliable in absolute terms, and much more so than those of the benchmark models. Focal maintains almost in-sample accuracy, showing that the in-sample accuracy was not an artefact of overfitting.*

Note that the reliability of counterfactual predictions is robust to limiting the knowledge about the target environment, that is, even precision (degree 2) and choice bias (degree 3) are predicted accurately, and thus represent robust facets of behavior. In applications, neither precision nor choice bias therefore need to be known for the target environment.

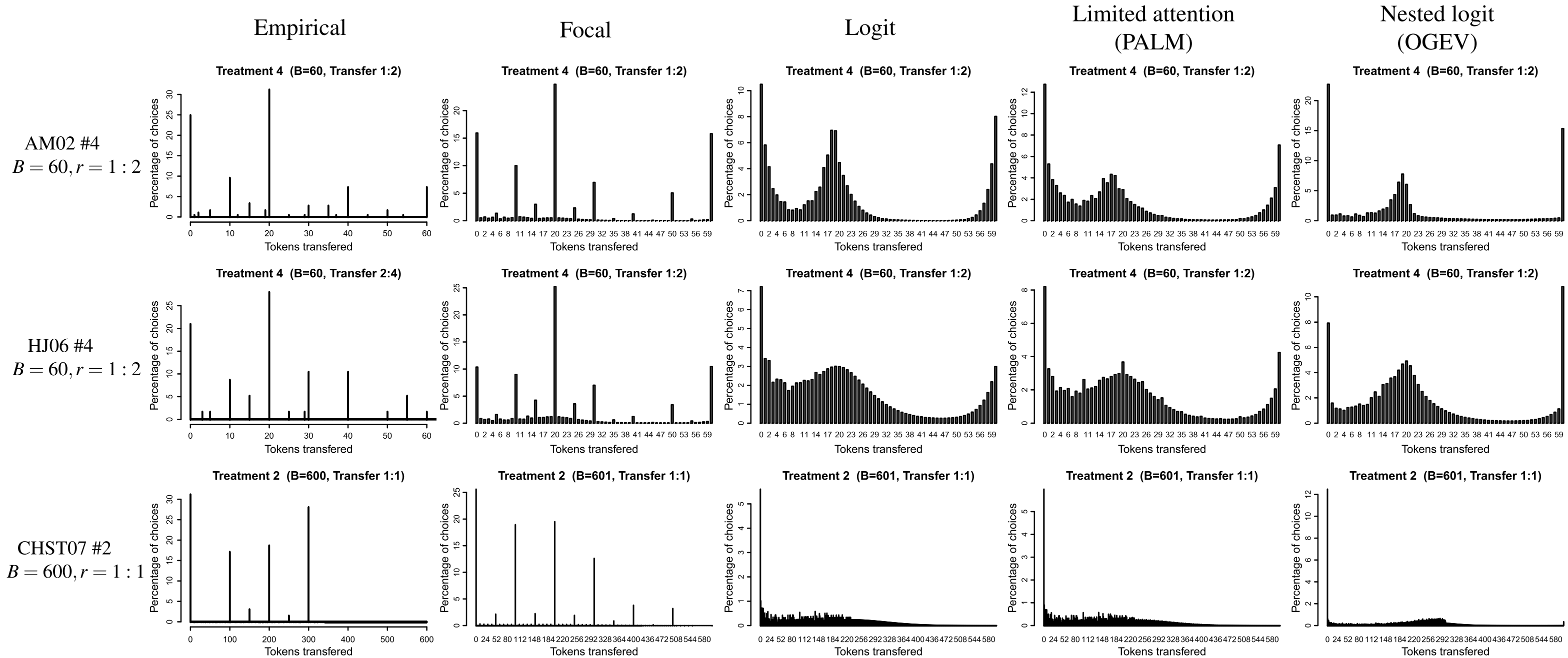

FIGURE 3. Impact of controlling for round-number effects in capturing manual choice. *Note*: These plots illustrate the predictions according to the four choice models, next to the empirical distribution, in representative treatments after fitting the model parameters to all treatments from the respective experiments. 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.

Consistency of estimates I evaluate the consistency of estimates between experiments in likelihood ratio tests, by comparing likelihoods in-sample and out-of-sample. Consistency is violated if in-sample and out-of-sample differ significantly. Obviously, estimates could be inconsistent simply due to differences in the subject pools, but in this case, all models will detect inconsistency. Note that consistency has been argued to be negatively correlated with in-sample accuracy, since a model that does not fit well in-sample may still be particularly robust due to being "simpler" (Hey, Lotito, and Maffioletti (2010)). This may put Focal at a disadvantage.

Again, I distinguish the three degrees. Consistency to the first degree evaluates the consistency of preference estimates in isolation, by relating the in-sample BIC to the out-of-sample BICs obtained from predictions to the first degree,

$$\Delta BIC_o^{\text{pred-1}} = \sum_{o' \neq o} \left[BIC_{o'|o}^{\text{pred-1}} - BIC_{o'}^{\text{descr}}\right]. \tag{12}$$

That is, I use preference estimates (distributions of α, β) of data set o, and given these parameters I estimate precision and choice bias (distributions of λ, κ) on any other data set o', in order to obtain a full set of predictions to the first degree based on preference estimates from o. Next, I take the differences of the resulting out-of-sample BICs to the respective in-sample BIC for each o' and aggregate, thus evaluating the consistency of the preference estimates (in terms of log-likelihood differences) obtained from data set o across other data sets. Consistency to the second degree uses out-of-sample preference and precision estimates, that is, the BICs of predictions to the second degree,

$$\Delta BIC_o^{\text{pred-2}} = \sum_{o' \neq o} \left[BIC_{o'|o}^{\text{pred-2}} - BIC_{o'}^{\text{descr}}\right] \tag{13}$$

and thus evaluates the joint consistency of preference and precision estimates across samples. Consistency to the third degree uses the out-of-sample estimates for all behavioral parameters in the prediction stage,

$$\Delta BIC_o^{\text{pred-3}} = \sum_{o' \notin \{o, \text{FKM07}\}} \left[BIC_{o'|o}^{\text{pred-3}} - LL_{o'}^{\text{descr}}\right]. \tag{14}$$

Note that for each data set we only have two sets of predictions to the third degree, namely the two sets based on estimates from the other two manual DG experiments, since we are focusing on manual experiments when evaluating predictions fully out-of-sample.

The third panel of Table 2 provides the results, listing the differences ΔBIC between in-sample and out-of-sample and using asterisks to indicate significance of differences. In total, Focal violates consistency at the Bonferroni level of 0.005 in 1/8 cases, while logit, PALM and OGEV yield estimates that violate consistency in 7/8 cases each. I also evaluate "relative consistency" by means of the relation signs in Table 2. A model is 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. 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.[5]

Result 3. *Preference estimation using the Focal model is consistent across large experiments (at least eight observations per subject). Estimates from the small experiment (CHST07) differ weakly significantly, indicating weak identification. Estimates obtained using the benchmark models are inconsistent across all experiments.*

Combined, the counterfactual predictions show that estimates from other experiments allow to predict behavior in CHST07 accurately, close to achieving in-sample accuracy, while estimate consistency shows that the estimates from CHST07 are not suitable to predict behavior in the other experiments. Jointly, this attests weak identification of preferences based on the CHST07 data (with only two observations per subject), as opposed to attesting differences between subject pools. That is, reliable preference measurement is possible with adequate choice models (such as Focal) and a sufficiently large number of observations per subject (at least 8–10, as in AM02 and HJ06).

Comparative relevance Finally, let me evaluate the economic relevance of controlling for presentation effects, both in absolute terms and in relative terms compared to the relevance of controlling for subject heterogeneity. To this end, I aggregate the BIC measures defined above across all data sets,

$$\begin{aligned} BIC^{\text{descr}} &= \sum_{o \in O} LL_o^{\text{descr}}, \qquad BIC^{\text{pred-1}} = \sum_{o \in O} LL_o^{\text{pred-1}}, \\ BIC^{\text{pred-2}} &= \sum_{o \in O} LL_o^{\text{pred-2}} \end{aligned} \tag{15}$$

in order to get overall assessments of descriptive and predictive adequacy, and using these aggregate BIC I compute the corresponding pseudo-R^2 as before.

These aggregate pseudo-R^2 are reported in Table 3. I focus on the results for logit and Focal, as the other benchmark models are similar to logit. As for modeling subject heterogeneity, I distinguish four model families: (i) representative agent models (called "Repr Agent" in Table 3) where all subjects have identical parameters; models with heterogeneous preferences ("Het Prefs") allowing for heterogeneity in preferences (random α and β with distributions as above) while subjects are assumed to have homogeneous precision λ and choice bias κ; models allowing for heterogeneity in preferences and precision (random α, β, λ as above, called "Het Pref and Prec" in Table 3) while subjects are assumed to have homogeneous choice bias κ, and models allowing for full heterogeneity (called "Full Het") used in the previous subsection. As indicated, I further distinguish the aggregate pseudo-R^2 in-sample ("Accuracy in-sample"), and of predictions to the first and second degree ("Out-of-sample degree 1" and "Out-of-sample degree 2").

[5]As above, there is no robust ranking between logit, PALM and OGEV. Between these three models, every one of them is most consistent in one context and least consistent in another context.

TABLE 3. Relevance of controlling for round-number effects compared to relevance of controlling for subject heterogeneity (Pseudo-R^2, less is better).

	Extent of Controlling for Heterogeneity						
	Repr Agent		Het Prefs		Het Pref and Prec		Full Het
Accuracy in-sample (pseudo-R^2 based on BIC^{descr}, see equation (15))							
Logit	0.212	$<$ (0)	0.48	$<$ (0)	0.598	$\approx$ (0.778)	0.598
Focal	0.603	$<$ (0)	0.863	$<$ (0.009)	0.884	$\approx$ (0.83)	0.885
Out-of-sample degree 1 (prediction of preferences, based on $BIC^{\text{pred-1}}$)							
Logit	−0.091	$<$ (0)	0.452	$>$ (0.002)	0.399	$\approx$ (0.572)	0.401
Focal	0.564	$<$ (0)	0.738	$<$ (0)	0.838	$\approx$ (0.991)	0.838
Out-of-sample degree 2 (prediction of preferences and precision, based on $BIC^{\text{pred-2}}$)							
Logit	−0.048	$<$ (0)	0.395	$\approx$ (0.867)	0.398	$\approx$ (0.718)	0.401
Focal	0.26	$<$ (0)	0.777	$<$ (0)	0.838	$\approx$ (0.641)	0.84

Note: As discussed in the text, each panel provides 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 (multinomial logit and Focal). The underlying likelihood ratio tests follow Schennach and Wilhelm (2016), the relation signs indicate the direction of differences, and the p-values of the LR tests are printed underneath the relation signs.

In-sample, the representative agent logit model captures 21.2% of observed variance, and controlling for heterogeneity allows us to explain up to 59.8%, that is, an additional 38 percentage points. Controlling for focality has the same impact in isolation, raising the pseudo-R^2 to 60.3% in the representative agent model, and in addition to controlling for heterogeneity, it improves the R^2 still by 30 percentage points (up to a pseudo-R^2 of 88.5% in total). That is, focality and subject heterogeneity are quantitatively of similar relevance and fairly complementary in nature.

The results out-of-sample are qualitatively similar but quantitatively even stronger in demonstrating the relevance of controlling for both heterogeneity and focality. In predictions to both the first and the second degree, the representative agent logit model has negative validity (pseudo-R^2 of −9.1% and −4.8% out-of-sample), implying that representative-agent predictions provide at best uninformed guesses. This is intuitive, as the comparative statics depend on presentation, and if we calibrate a representative agent based on one experiment in Figure 2, then the predictions for other experiments, with differing contribution levels and comparative statics, must be off. Representative agent modeling seems to be fairly limited as far as giving (as in dictator games) is concerned.

The reliability of preference estimation (prediction to the first degree) improves by around 49 percentage points in terms of R^2 by controlling for heterogeneity, up to 40.1% in total, and by another 43 percentage points on top controlling for focality (up to 83.8%). The reliability of joint preferences and precision estimation (prediction to the second degree), that is, the distributions of (α, β, λ) across subjects, improves by 45 percentage points controlling for heterogeneity (up to again 40.1%) and controlling for focality adds another 44 percentage points on top (up to 84%).

Result 4. *Controlling for focality and controlling for heterogeneity are both highly relevant and largely complementary in reliable preference measurement.*

5. Concluding remarks

This paper introduces and tests a simple model allowing to capture presentation effects in stochastic choice. The idea that presentation 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, for example, Bernheim and Rangel (2007) and Salant and Rubinstein (2008), all focus on rational (nonstochastic) choice, and thus are of limited help in econometric analyses. This paper contributes to this literature by showing the striking importance, comparable to controlling for heterogeneity, and the actual possibility of controlling for presentation effects in experimental data. The focal choice adjust logit (Focal) model developed here is tested by reanalyzing seminal experiments on dictator games, the most widely analyzed experimental games, and the results confirm the impression of the introductory comments. Choice patterns and comparative statics strongly depend on presentation, and inadvertently, counterfactual predictions and utility analyses neglecting presentation are unreliable (Results 2 and 3). This implies that falsifying predictions or detecting behavioral differences, by analyzing experiments without controlling for presentation, risks being uninformative or misleading.

The econometric results explicitly show that we can and need to control for presentation when making ex ante predictions for experiments or policy interventions, and equally when evaluating behavioral models. The latter requires rich data sets, with multiple observations per subject to disentangle preferences, noise and presentation bias. Without such experimental designs, estimates are inconsistent (for all models) and convergence of say preference modeling appears unrealistic. In the econometric analysis, I assumed that the focality of options is linear in the "level of roundedness" observed in statistical analysis of survey responses. This can be generalized straightforwardly, but linearity ex post seems sufficient when analyzing dictator choice. Further, 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 simple reversions do not nullify ordering effects. Proposition 1 suggests an alternative approach, namely to arrange the options along a circle, which can be rotated, and the resulting equation system identifies utility and focality indices up to affine transformation. 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.

To conclude, the Focal model allows to control for the "focality" of options arising from the presentation of choice tasks. This is as critical for external validity as controlling for heterogeneity, as presentation affects even comparative statics in dictator games. The Focal model is widely applicable, as it directly generalizes logit, which is the workhorse model in analyses of both individual choice (as analyzed here) and strategic choice, including quantal response equilibrium (McKelvey and Palfrey (1995), McKelvey and Palfrey (1998)), cognitive hierarchy models (Camerer, Ho, and Chong (2004), Rogers, Palfrey, and Camerer (2009)), level-k models (Stahl and Wilson (1994), Costa-Gomes, Crawford, and Broseta (2001), Costa-Gomes and Weizsäcker (2008)) and noisy introspection (Goeree and Holt (2004)). In this sense, Focal is a versatile and promising model for behavioral analyses, besides being a rigorous framework to study presentation effects including nudging interventions. This opens up a wide range of interesting opportunities for analyses of behavior across experiments, as controlling for presentation enables joint analyses of multiple experiments, which in turn allows us to condense the information contained therein.

Appendix: Proof and technical definitions

A.1 *Proof of Proposition* 1

Fix $(\pi, \phi) \in \mathcal{C}$, $B \subset X$, and define $p_x := \log \Pr(x|\pi, \phi, X)$ for all $x \in B$. Further, define $u_x = u(\pi_x)$ and $v_x = v(\phi_x)$, for all x.

Step 1: There exists $a \in \mathbb{R}$ such that $p_x = \lambda u_x + \kappa \phi_x + a$ for all x.

If Pr has a Focal representation, then by $p_x = \log \Pr(x|\pi, \phi, X)$ that

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

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

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

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

$$\frac{\exp\{\lambda u_x + \kappa v_x\}}{\sum_{x' \in B} \exp\{\lambda u_{x'} + \kappa v_{x'}\}} = \frac{\exp\{\lambda u_x + \kappa v_x + a(x)\}}{\sum_{x' \in B} \exp\{\lambda u_{x'} + \kappa v_{x'} + a(x')\}}$$

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

implying $a(x)/a(x') = 1$ for all $x, x' \in X$. Thus, there exists $a \in \mathbb{R}$ such that $p_x = \lambda u_x + \kappa v_x + a$ for all x.

Step 2: Fix $B \in P(X)$ and any context (π, ϕ). Initially, assume that the second context $\tilde{\pi}$ is obtained by the "simple rotation" of π toward $\tilde{\pi}$ where

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

We show that observations from the two contexts (π, ϕ) and $(\tilde{\pi}, \phi)$ suffice to identify (π, ϕ) up to affine transformation.

Let $n = |B|$. By Theorem 1, based on the choice probabilities observing in (π, ϕ) and $(\tilde{\pi}, \phi)$, we obtain the equation system:

$$p_1 = a + \lambda u_1 + \kappa v_1 \tag{1a}$$
$$\tilde{p}_1 = \tilde{a} + \lambda u_n + \kappa v_1 \tag{1b}$$
$$p_2 = a + \lambda u_2 + \kappa v_2 \tag{2a}$$
$$\tilde{p}_2 = \tilde{a} + \lambda u_1 + \kappa v_2 \tag{2b}$$
$$\vdots$$
$$p_n = a + \lambda u_n + \kappa v_n \tag{na}$$
$$\tilde{p}_n = \tilde{a} + \lambda u_{n-1} + \kappa v_n, , \tag{nb}$$

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

We demonstrate this directly by solving the equation system. 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:

$$p_1 - \tilde{p}_1 = a - \tilde{a} + \lambda u_1 - \lambda u_n \tag{1c}$$
$$p_1 - \tilde{p}_2 = a - \tilde{a} + \lambda v_1 - \lambda v_2 \tag{1d}$$
$$p_2 - \tilde{p}_2 = a - \tilde{a} + \lambda u_2 - \lambda u_1 \tag{2c}$$
$$p_2 - \tilde{p}_3 = a - \tilde{a} + \lambda v_2 - \lambda v_3 \tag{2d}$$
$$\vdots$$
$$p_n - \tilde{p}_n = a - \tilde{a} + \lambda u_n - \lambda u_{n-1} \tag{nc}$$
$$p_n - \tilde{p}_1 = a - \tilde{a} + \lambda v_n - \lambda v_1 \tag{nd}$$

Second, set $v_1 = 0$ and define equations $(1e) = (1d)$, $(2e) = (1d) + (2d)$, $(3e) = (2e) + (3d)$, $\dots$, $(10e) = (9e) + (10d)$, as well as constants $d_1 = p_1 - \tilde{p}_2$, $d_2 = p_2 - \tilde{p}_3$ and so on. This yields

$$d_1 = a - \tilde{a} - \lambda v_2, \tag{1e}$$
$$d_1 + d_2 = 2(a - \tilde{a}) - \lambda v_3, \tag{2e}$$
$$d_1 + d_2 + d_3 = 3(a - \tilde{a}) - \lambda v_4, \tag{3e}$$
$$\vdots \tag{16}$$
$$d_1 + \dots + d_{n-1} = (n-1)(a - \tilde{a}) - \lambda v_n, \tag{(n-1)e}$$
$$d_1 + \dots + d_n = n(a - \tilde{a}) - \lambda v_1. \tag{ne}$$

Given $v_1 = 0$, equation (ne) defines $(a - \tilde{a})$, and using this, $\lambda v_2, \dots, \lambda v_{10}$ are defined from equations (1e)–((n-1)e).

Finally, set $u_1 = 0$, which implies that λu_n 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_{n-1} which is identified from (9c).

Step 3: If $\tilde{\pi}$ is any circular permutation of π, then relabeling options transforms the circular permutation into a simple rotation as defined in Step 2. Hence, the claim holds true for all circular permutations. □

A.2 *A simple axiomatic foundation*

Based on Breitmoser (2020), it is straightforward to provide a simple axiomatic foundation of the Focal model. The starting point that I use in this analysis is Gul and Pesendorfer (2001), who establish that a "rational" DM (with a preferences for commitment) maximizes $u + v$. The question I ask is, if we take the existence of u and v as given, in which conditions does "injecting" stochastic choice imply the Focal model. Formally, I inject stochastic choice by requiring positivity of choice probabilities, then translate $\max u + v$ into invariance statements about choice probabilities, and finally solve for the functional form that choice probabilities may take if they are compatible with these invariance statements. In order to reflect that we now take u and v as given, and simply ask about the functional form relating choice probabilities with u and v, I now denote choice probabilities as $\Pr(x|u, v, B)$. Relatedly, each pair (u, v) characterizes a DM, and the set of all conceivable DM is denoted $\mathcal{D}$. I write $u_x = u(x)$ and $v_x = v(x)$ for all $x \in X$.

The invariance statements are that both dimensions u and v matter (essentialness, otherwise the functional form is not unique), that choice satisfies independence of irrelevant alternatives (IIA), that choice depends solely on u and v ("completeness"), and that it is invariant to translation of u and v ("narrow bracketing" and "relative focality"), that is, that DMs with u or v differing only by translation choose options with equal probabilities. Note that all of these statements are compatible with $\max u + v$, and in this sense, they (partially) characterize $\max u + v$. As indicated, in addition, we assume positivity of choice probabilities.

Assumption 1 (Axioms). *For all $(u, v) \in \mathcal{D}$, all $B \in P(X)$, all $x \in B$ and all $r \in \mathbb{R}$,*

1. *Essentialness*: $u_x \neq u_y$ *and* $v_x = v_y$ *implies* $\Pr(x|u, v, B) \neq \Pr(y|u, v, B)$, *and* $u_x = u_y$ *and* $v_x \neq v_y$ *implies* $\Pr(x|u, v, B) \neq \Pr(y|u, v, B)$,
2. *Positivity*: $$\Pr(x|u, v, B) > 0,$$
3. *IIA*: *for all* $B' \in P(X)$ *and all* $x, y \in B \cap B'$, $$\frac{\Pr(x|u,v,B)}{\Pr(y|u,v,B)} = \frac{\Pr(x|u,v,B')}{\Pr(y|u,v,B')},$$
4. *Narrow bracketing*: $$\Pr(x|u, v, B) = \Pr(x|u + r, v, B),$$
5. *Relative focality*: $$\Pr(x|u, v, B) = \Pr(x|u, v + r, B),$$
6. *Completeness*: *for all* $(u, v), (u', v') \in \mathcal{D}$, $B = \{x, y\}$, *and* $B' = \{x', y'\}$, $$(u_x, v_x) = (u'_{x'}, v'_{x'}) \text{ and } (u_y, v_y) = (u'_{y'}, v'_{y'}) \Rightarrow \frac{\Pr(x|u,v,B)}{\Pr(y|u,v,B)} = \frac{\Pr(x'|u',v',B')}{\Pr(y'|u',v',B')}.$$

We solve for the functional form that choice probabilities may take, given they are compatible with these statements, assuming the following assumptions are satisfied. First, we need to make sure that we can manipulate u and v independently (Richness). Second, for any conceivable DM (u, v), we need to ensure that DMs for all affine transformations of u and v are equally conceivable, based on which I show that DMs differing only by affine transformations actually are indistinguishable.[6] Finally, to guarantee uniqueness, the set of possible options X as well as the images of u and v are convex subsets of $\mathbb{R}$.

Assumption 2 (Richness). 1. *Choice tasks*: *For all* $(u, v) \in \mathcal{D}$, $(a_u + b_u \cdot u, a_v + b_v \cdot v) \in \mathcal{D}$ *for all* $(a_u, b_u), (a_v, b_v) \in \mathbb{R}^2$.

2. *Richness*: *there exist* (u, v) *and* (x, y) *such that* $u_x \neq u_y$ *and* $v_x = v_y$, *and there exist* (u, v) *and* (x, y) *such that* $u_x = u_y$ *and* $v_x \neq v_y$.

3. *Convexity*: X *and the images* $u[X] = \{u_x | x \in X\}$, $v[X] = \{v_x | x \in X\}$ *are convex, bounded, nonsingleton subsets of* $\mathbb{R}$.

The following result establishes that the above invariance statements are satisfied if and only if choice probabilities have a Focal representation. Considering that the invariance statements simply characterize $\max u + v$, and that the only other assumption is positivity, this suggests that the Focal model is a natural formulation of $\max u + v$ if we seek to allow for stochastic choice.

Theorem 1. *Given Assumption* 2, *the following two statements are equivalent*:

1. *Choice profile* Pr *satisfies Axioms* 1–6

2. *Choice profile* Pr *has a Focal choice adjusted logit* (*Focal*) *representation*: *there exist unique* $(\lambda, \kappa) \in \mathbb{R}^2$ *such that, for all* $x \in B \in P(X)$ *and all* $(u, v) \in \mathcal{D}$,

$$\Pr(x|u, v, B) = \frac{\exp\{\lambda \cdot u_x + \kappa \cdot v_x\}}{\sum_{x' \in B} \exp\{\lambda \cdot u_{x'} + \kappa \cdot v_{x'}\}}. \tag{17}$$

Since a formal proof of a slightly more general statement allowing for an arbitrary number of dimensions n is provided in Breitmoser (2020), I will resort to outline the key intuition here. 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 DM (u, v) separately, propensities may be DM dependent, implying

$$\Pr(x|u, v, B) = \frac{V(x, y|u, v)}{\sum_{x' \in B} V(x', y|u, v)} \quad \text{with } V(x, y|u, v) = f_{u,v}(u_x, u_y, v_x, v_y, x, y).$$

[6] I 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 $x \in X$.

Any $y \in X$ may serve as benchmark option, and by convexity of X, the reference to the benchmark can be dropped, that is, $V(x, y|u, v) = \tilde{f}_{u,v}(u_x, v_x, x)$ for some function $\tilde{f}$. By Axiom 6, 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 $\tilde{V}_{u,v}$ exists such that $\tilde{V}_{u,v}(u_x, v_x) = \tilde{f}_{u,v}(u_x, v_x, x)$ for all $x \in X$, and all (u, v). Given this characterization of propensities, narrow bracketing and focality imply

$$\frac{\tilde{V}_{u,v}(u_x, v_x)}{\tilde{V}_{u,v}(u_{x'}, v_{x'})} = \frac{\tilde{V}_{u+r_u,v+r_v}(u_x + r_u, v_x + r_v)}{\tilde{V}_{u+r_u,v+r_v}(u_{x'} + r_u, v_{x'} + r_v)} \quad \text{for all } x, x' \in X \text{ and } r_u, r_v \in \mathbb{R}, \tag{18}$$

which in turn implies $\tilde{V}_{u+r_u,v+r_v}(u_x + r_u, v_x + r_v) = \tilde{V}_{u,v}(u_x, v_x) \cdot g(r_u, r_v)$ for some function g. Due to the DM dependence of $\tilde{V}$, $\tilde{V}_{u+r_u,v+r_v} \neq \tilde{V}_{u,v}$ is possible, but "completeness" restricts DM dependence by allowing the functional equation to be expressed as $h(u_x + r_u, v_x + r_v) = h(u_x, v_x) \cdot g(r_u, r_v)$ for some function $h : \mathbb{R}^2 \to \mathbb{R}$. The main technical difficulty is that h is not necessarily differentiable.[7] By positivity, the logarithmic transformation $\tilde{h} = \log h$ and $\tilde{g} = \log g$ is admissible, which yields the Pexider functional equation $\tilde{h}(u_x + r_u, v_x + r_v) = \tilde{h}(u_x, v_x) + \tilde{g}(r_u, r_v)$, and by their relation to probabilities, $\tilde{h}$ is bounded from above for all values in the images of u and v (for any DM),[8] which each have positive length by "convexity" (Assumption 2). This implies that all solutions of $\tilde{h}$ are linear in u_x and v_x, and the essentialness of u and v implies that the respective coefficients (λ, κ) are unique. Thus, $\tilde{h}(u_x, v_x) = \lambda u_x + \kappa v_x + c_x$ for all x, with unique $\lambda, \kappa \in \mathbb{R}$ and $c : X \to \mathbb{R}$ (Aczél and Dhombres (1989)). Using $\tilde{V}_{u,v} = \exp \tilde{h}$,

$$\Pr(x|u, v, B) = \frac{\exp\{\lambda \cdot u_x + \kappa \cdot v_x + c_x\}}{\sum_{x' \in B} \exp\{\lambda \cdot u_{x'} + \kappa \cdot v_{x'} + c_{x'}\}}. \tag{19}$$

Completeness implies that c_x is constant in x and cancels out, yielding the Focal representation. The detailed proof is given in Breitmoser (2020).

References

Aczél, J. and J. G. Dhombres (1989), *Functional Equations in Several Variables*. Cambridge University Press. [277]

Andersen, S., G. Harrison, M. Lau, and E. Rutström (2008), "Eliciting risk and time preferences." *Econometrica*, 76 (3), 583–618. [262]

Andreoni, J. and J. Miller (2002), "Giving according to GARP: An experimental test of the consistency of preferences for altruism." *Econometrica*, 70 (2), 737–753. [252, 260]

[7] For purpose of illustration, consider the simple case that propensities $\tilde{V}$ are DM independent, differentiable in u_x and independent of v_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 $c \in \mathbb{R}$.

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

Battistin, E., R. Miniaci, and G. Weber (2003), "What do we learn from recall consumption data?" *Journal of Human Resources*, 38 (2), 354–385. [261]

Bellemare, C., S. Kröger, and A. van Soest (2008), "Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities." *Econometrica*, 76 (4), 815–839. [262]

Bernheim, B. D. and A. Rangel (2007), "Toward choice-theoretic foundations for behavioral welfare economics." *American Economic Review*, 97 (2), 464–470. [272]

Bordalo, P., N. Gennaioli, and A. Shleifer (2012), "Salience theory of choice under risk." *Quarterly Journal of Economics*, 127 (3), 1243–1285. [254]

Breitmoser, Y. (2020), "An axiomatic foundation of conditional logit." *Economic Theory*. [257, 275, 276, 277]

Breitmoser, Y. (2021), "Supplement to 'Controlling for presentation effects in choice'." *Quantitative Economics Supplemental Material*, 12, https://doi.org/10.3982/QE1050. [254]

Camerer, C., T. Ho, and J. Chong (2004), "A cognitive hierarchy model of games." *Quarterly Journal of Economics*, 119 (3), 861–898. [260, 273]

Camerer, C. and T. H. Ho (1999), "Experience-weighted attraction learning in normal form games." *Econometrica*, 67 (4), 827–874. [260]

Cappelen, A., A. Hole, E. Sørensen, and B. Tungodden (2007), "The pluralism of fairness ideals: An experimental approach." *American Economic Review*, 97 (3), 818–827. [252, 260, 261, 262]

Costa-Gomes, M., V. Crawford, and B. Broseta (2001), "Cognition and behavior in normal-form games: An experimental study." *Econometrica*, 69 (5), 1193–1235. [273]

Costa-Gomes, M. and G. Weizsäcker (2008), "Stated beliefs and play in normal-form games." *Review of Economic Studies*, 75 (3), 729–762. [273]

Covey, J. and R. Smith (2006), "How common is the 'prominence effect'?" *Additional evidence to Whynes et al. Health economics*, 15 (2), 205–210. [261]

De Bruyn, A. and G. E. Bolton (2008), "Estimating the influence of fairness on bargaining behavior." *Management Science*, 54 (10), 1774–1791. [260]

Dean, M. L. (1980), "Presentation order effects in product taste tests." *The Journal of psychology*, 105 (1), 107–110. [251]

Dinner, I., E. J. Johnson, D. G. Goldstein, and K. Liu (2011), "Partitioning default effects: Why people choose not to choose." *Journal of Experimental Psychology: Applied*, 17 (4), 332. [251]

Echenique, F., K. Saito, and G. Tserenjigmid (2018), "The perception-adjusted Luce model." *Mathematical Social Sciences*, 93, 67–76. [263, 264]

Feenberg, D., I. Ganguli, P. Gaule, and J. Gruber (2017), "It's good to be first: Order bias in reading and citing NBER working papers." *Review of Economics and Statistics*, 99 (1), 32–39. [251]

Fisman, R., S. Kariv, and D. Markovits (2007), "Individual preferences for giving." *American Economic Review*, 97 (5), 1858–1876. [252, 260, 261]

Goeree, J. and C. Holt (2004), "A model of noisy introspection." *Games and Economic Behavior*, 46 (2), 365–382. [273]

Gul, F. and W. Pesendorfer (2001), "Temptation and self-control." *Econometrica*, 69 (6), 1403–1435. [254, 256, 275]

Gul, F. and W. Pesendorfer (2006), "The simple theory of temptation and self-control." [257]

Halevy, Y., D. Persitz, and L. Zrill (2018), "Parametric recoverability of preferences." *Journal of Political Economy*, 126 (4), 1558–1593. [261]

Harless, D. and C. Camerer (1994), "The predictive utility of generalized expected utility theories." *Econometrica*, 1251–1289. [260]

Harrison, G. W. and L. T. Johnson (2006), "Identifying altruism in the laboratory." In *Experiments Investigating Fundraising and Charitable Contributors* (R. M. Isaac and D. D. Davis, eds.), Research in Experimental Economics, Vol. 11, 177–223, Emerald Group Publishing Limited. [252, 260]

Harrison, G. W., J. A. List, and C. Towe (2007), "Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion." *Econometrica*, 75 (2), 433–458. [262]

Heitjan, D. F. and D. B. Rubin (1991), "Ignorability and coarse data." *Annals of Statistics*, 2244–2253. [251]

Hey, J., G. Lotito, and A. Maffioletti (2010), "The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity." *Journal of risk and uncertainty*, 41 (2), 81–111. [260, 269]

Kreps, D. M. (1990), *Game Theory and Economic Modelling*. Oxford University Press. [256]

Lacetera, N., D. G. Pope, and J. R. Sydnor (2012), "Heuristic thinking and limited attention in the car market." *American Economic Review*, 102 (5), 2206–2236. [251]

Luce, R. (1959), *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York, NY. [276]

Manski, C. F. and F. Molinari (2010), "Rounding probabilistic expectations in surveys." *Journal of Business & Economic Statistics*, 28 (2), 219–231. [251]

Manzini, P. and M. Mariotti (2014), "Stochastic choice and consideration sets." *Econometrica*, 82 (3), 1153–1176. [263, 264]

Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012), "Revealed attention." *American Economic Review*, 102 (5), 2183–2205. [263]

Matejka, F. and A. McKay (2015), "Rational inattention to discrete choices: A new foundation for the multinomial logit model." *American Economic Review*, 105 (1), 272–298. [257]

McFadden, D. (1974), "Conditional logit analysis of qualitative choice models." In *Frontiers of Econometrics* (P. Zarembka, ed.), 105–142, Academic Press, New York, NY. [276]

McFadden, D. (1976), "Quantal choice analysis: A survey." *Annals of Economic and Social Measurement*, 5 (4), 363–390. [263]

McKelvey, R. D. and T. R. Palfrey (1995), "Quantal response equilibria for normal form games." *Games and Economic Behavior*, 10 (1), 6–38. [273]

McKelvey, R. D. and T. R. Palfrey (1998), "Quantal response equilibria for extensive form games." *Experimental Economics*, 1 (1), 9–41. [273]

McKenzie, C. R., M. J. Liersch, and S. R. Finkelstein (2006), "Recommendations implicit in policy defaults." *Psychological Science*, 17 (5), 414–420. [251]

Miller, J. M. and J. A. Krosnick (1998), "The impact of candidate name order on election outcomes." *Public Opinion Quarterly*, 62, 291–330. [251]

Nagelkerke, N. J. (1991), "A note on a general definition of the coefficient of determination." *Biometrika*, 78 (3), 691–692. [264]

Poltrock, S. E. and D. R. Schwartz (1984), "Comparative judgments of multidigit numbers." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10 (1), 32. [251]

Powell, M. (2006), "The newuoa software for unconstrained optimization without derivatives." *Large-Scale Nonlinear Optimization*, 255–297. [262]

Rogers, B., T. R. Palfrey, and C. Camerer (2009), "Heterogeneous quantal response equilibrium and cognitive hierarchies." *Journal of Economic Theory*, 144 (4), 1440–1467. [273]

Rubinstein, A. and Y. Salant (2006), "A model of choice from lists." *Theoretical Economics*, 1 (1), 3–17. [272]

Salant, Y. and A. Rubinstein (2008), "(a, f): Choice with frames." *The Review of Economic Studies*, 75 (4), 1287–1296. [272]

Schelling, T. C. (1960), *The Strategy of Conflict*. Harvard University Press. [254, 256]

Schennach, S. and D. Wilhelm (2016), "A simple parametric model selection test." *Journal of the American Statistical Association*. [262, 266, 271]

Schwarz, G. (1978), "Estimating the dimension of a model." *Annals of Statistics*, 6 (2), 461–464. [264]

Sims, C. A. (2003), "Implications of rational inattention." *Journal of Monetary Economics*, 50 (3), 665–690. [257]

Small, K. (1987), "A discrete choice model for ordered alternatives." *Econometrica*, 55 (2), 409–424. [263]

Stahl, D. O. and P. W. Wilson (1994), "Experimental evidence on players' models of other players." *Journal of Economic Behavior and Organization*, 25 (3), 309–327. [273]

Sugden, R. (1995), "A theory of focal points." *The Economic Journal*, 105 (430), 533–550. [256]

Train, K. (2003), *Discrete Choice Methods With Simulation*. Cambridge University Press. [262]

Whynes, D., Z. Philips, and E. Frew (2005), "Think of a number… any number?" *Health Economics*, 14 (11), 1191–1195. [261]

Wilcox, N. (2008), "Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison." In *Research in Experimental Economics* (J. C. Cox and G. W. Harrison, eds.), Risk Aversion in Experiments, Vol. 12, 197–292, Emerald Group Publishing Limited. [260, 262]

Co-editor Christopher Taber handled this manuscript.

Manuscript received 11 January, 2018; final version accepted 17 September, 2020; available online 24 September, 2020.