Diagnosing pervasive issues with parameter estimation

Sabina J. Sloman (SSLOMAN@Andrew.Cmu.Edu)
Stephen B. Broomell (BROOMELL@Cmu.Edu)
Timothy Kusuma (TKUSUMA@Andrew.Cmu.Edu)
Department of Social and Decision Sciences, Carnegie Mellon University, 5000 Forbes Avenue
Pittsburgh, PA 15213 USA

Abstract

We explore structural issues with parameter estimation for non-linear cognitive models: Some parameter values are easier to recover than others, and the recoverability of different parameters interacts in systematic ways. We propose methods for researchers to anticipate and visualize and these issues, and the systematic ways they differ across experimental designs. Our approach consists of assessing how changes in parameter values translate into changes in behavioral predictions, and develop measurements of the relative responsiveness of predictions to parameter values. We demonstrate application of our approach to cumulative prospect theory (CPT), a widely-used model of risky decision-making.

Keywords: decision-making; prospect theory; parameter estimation

Experimental design plays a large role in the precision of the parameter estimates of models of decision-making and other cognitive processes (Broomell & Bhatia, 2014; Broomell, Sloman, Blaha, & Chelen, 2019; Navarro, Pitt, & Myung, 2004; Stewart, Canic, & Mullett, 2019; Toubia, Johnson, Evgeniou, & Delquié, 2013). Understanding the implications of the parameter estimates recovered from a dataset requires an understanding of several aspects of the model-stimulus relationship, including the diagnosticity of the experimental stimuli (Broomell et al., 2019). In order to facilitate this understanding, Broomell et al. (2019) suggest researchers report and discuss the process of stimulus selection in model comparison and parameter recovery studies.

Broomell and Bhatia (2014) analyze the parameter discrimination of decision sets used to estimate the parameters of cumulative prospect theory. While their work formalizes the discrimination power of a decision set, their measures do not account for structural differences in the ease with which parameter values that fall within different ranges can be recovered. In the present work, we conceptualize the parameter discrimination of a stimulus set as a *function* of the underlying parameter value rather than a static quantity. Like Broomell and Bhatia (2014), we focus our analysis on cumulative prospect theory. This paper exposes structural asymmetries in the recoverability of the parameters of CPT, and demonstrates the application of methods that can anticipate how an experimental design will exacerbate or mitigate these asymmetries. The aim of our work is to begin to develop a set of methods researchers can use to explore, visualize and document the strengths and weaknesses of certain experimental designs for parameter estimation.

A related body of work has focused on the development of methods in optimal experimental design for model comparison. Design optimization (Myung & Pitt, 2009) and adaptive design optimization (Cavagnaro, Myung, Pitt, & Kujala, 2010; Kim, Pitt, Lu, Steyvers, & Myung, 2014) are a collection of methods for designing experiments with maximum power to discriminate between competing models. Similar methods that optimize for parameter recovery (e.g. Toubia et al. (2013)) dynamically adjust the stimuli presented to a participant to maximize the precision of parameter estimates. While we demonstrate a method that achieves some success at enhancing the power of stimulus sets to estimate certain parameter values, this is intended primarily as a proof of concept. Our focus is not on optimization, but on anticipation of and transparency about the effects of the experimental design in parameter recovery studies.

In the results we present below, we show that designs with more power to estimate some parameter values have less power to estimate others. When participants exhibit large individual differences in the parameter values that best describe them, this may make it difficult to design an experiment that can effectively optimize for discrimination across the range of plausible parameter values. Dynamic approaches like those developed by Toubia et al. (2013) could bypass the trade-offs this implies. However, these methods cannot be used retrospectively to help researchers analyze archival datasets or to implement a robust static experimental design. Our work can facilitate the diagnosis and documentation of the strengths and weaknesses of a stimulus set: which parameter values it will have power to detect, and which it will not.

Cumulative prospect theory

Cumulative prospect theory (CPT) is a widely-used model of risky choice (Tversky & Kahneman, 1992). Parameter estimates from CPT have been used to understand individual- and group-level differences in psychological constructs like risk aversion, loss aversion and probability weighting (Scheibehenne & Pachur, 2015).

In the field of judgment and decision-making, risky decision-making is typically operationalized as a two-alternative forced choice task between probabilistic gambles. As an example, imagine being offered a choice between a 37% chance of winning \$10, or a 45% chance of winning \$2. Participants in risky decision-making tasks are presented

©2020 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

with dozens of such artificial choices, where both the outcome amounts and probability of winning these outcomes are varied.

Model formulation

Below we present a simplified formulation of CPT. For the complete model specification, interested readers should consult Tversky and Kahneman (1992), Glöckner and Pachur (2012) or Broomell and Bhatia (2014).

According to CPT, choice is predicted by combining the *subjective value* of possible outcomes X_j, $v(X_j)$, with the *weighted probability* of receiving outcome X_j, $w(p_j)$. In particular,

$$v(X_j) = X_j^{\alpha} \quad \text{and} \quad w(p_j) = e^{-(-ln(p_j))^{\gamma}} \tag{1}$$

The value of an entire gamble G_i is a combination of its constituent outcomes, which we denote,

$$V(G_i) = \sum v(y_j) w(p_j) \tag{2}$$

and the probability $P(G_i)$ of selecting G_i is a monotonic function of how much higher the value of G_i is than the alternative[1]:

$$P(G_1) = \frac{1}{1 + e^{-\varepsilon(V(G_1) - V(G_2))}} \tag{3}$$

The parameter α determines a decision-maker's (DM) degree of diminishing sensitivity to increased outcomes. As α decreases, larger outcomes generate proportionately less subjective value. The parameter γ determines the curvature of the probability weighting function. Smaller values of γ result in more over-weighting of small probabilities and under-weighting of high probabilities. Finally, ε encodes the determinism of the DMs' choices.

Parameter estimation

CPT parameters are often estimated from behavioral data using maximum likelihood estimation (Broomell & Bhatia, 2014; Stott, 2006). Maximum likelihood techniques attempt to find the values of a model's free parameters that make the observed data seem very probable. In practice, this objective is usually achieved by generating stochastic samples of candidate parameter values, and substituting these parameter values into the model formulation to generate a prediction. The likelihood is a measure of how closely this prediction matches the observed data. By computing the likelihood of multiple sets of candidate parameter values, the algorithm hones in on the values of the free parameters that *maximize the likelihood* of the observed data.

Imagine a researcher trying to infer the temperature without access to a direct measurement device like a thermometer. A maximum likelihood approach would compare different plausible values of the relevant parameter—degrees Farenheit—and select the one that maximizes the likelihood of a particular observation (or set of observations). For example, the researcher could base their guess on an observation that it's snowing. An initial guess of 60°F would imply that this observation is very unlikely. On the other hand, any guess between 15°and 35°F would imply that the observed weather was very likely. Therefore, the researcher's final estimate would be somewhere in that range.

In the case of CPT, the model's free parameters are α, γ and ε. As in the example above, the observed data are binary observations: the choice participants made between one of two risky gambles. The model's prediction is $P(G_1)$, the probability that a DM will select gamble 1.

Structural asymmetries in parameter estimation

When tweaking a parameter value translates into detectable changes in the outcomes predicted by a model, maximum likelihood routines can be highly efficient and generate accurate estimates. However, many theoretically interesting cognitive models are non-linear, meaning the degree of change in the predicted outcomes depends on both the underlying parameter values and on where in the domain the change is applied.

To continue with the example above, imagine the researcher again trying to infer the average temperature on the basis of several observations of whether or not it is snowing. If the true, generating parameter value is 30°F, when the researcher collects their observations, they will probably find that it is snowing much of the time. As explained above, the candidate parameter value 60°F would imply that these observations were very unlikely, while a guess closer to the true value of 30°F would imply that these observations were very likely. This design has a fair amount of power to identify parameters within a critical range.

However, if the true, generating parameter value *is* 60°F, when the researcher collects their observations, they probably won't observe any snow. While they can determine that the temperature is probably not between 15°and 35°F, they'll be unable to determine if it is 40°F, 60°F or 100°F, since all these parameter values predict the observed sequence. Relying solely on the binary outcome of snow will mean that there are certain regions of the parameter space that will be much more difficult to identify.

The same intuition applies to asymmetries in the identifiability of the parameter values of models like CPT. The estimates of CPT parameters are similarly based on binary observations—whether or not the DM selected to play one of two gambles—and the context in which those observations occurred. Some parameter values will lead to observations that systematically provide less information than others. But unlike the relationship between temperature and snow, the relationship between a CPT parameter value and the relevant binary observation is rarely obvious or transparent.

In the sections below, we identify some of the structural asymmetries in the estimation of CPT parameters. We generate synthetic data sets from simulated CPT DMs, and analyze

[1] We use the Prelec probability weighting function (Prelec, 1998) and a logit form of the choice function, but see Stott (2006) for a review of other forms of these functions.

differences in the recoverability of different parameter settings. Finally, we develop a conceptual framework to visualize and better understand how much information we can gain from a given experimental design, and show that this framework can both explain and anticipate the structural asymmetries that arise from different designs.

Stimulus selection

To generate these data sets, we first needed to select stimuli, or pairs of gambles that the artificial DMs would choose between. We generated 1,000 random gambles according to the procedure described in Erev, Roth, Slonim, and Barron (2002). We randomly sampled probabilities and outcome values to compose choices between gambles such that gamble i yields some value X_i with probability p_i, and nothing with probability $1-p_i$. Gambles i and j were paired together under the constraint that if $X_i > X_j$, then $p_j > p_i$ (and vice versa).

One-parameter model

For the sake of interpretability, we first explore a one-parameter version of CPT. Specifically, we constrain both γ and ε to be .5 and allow only α to vary freely.

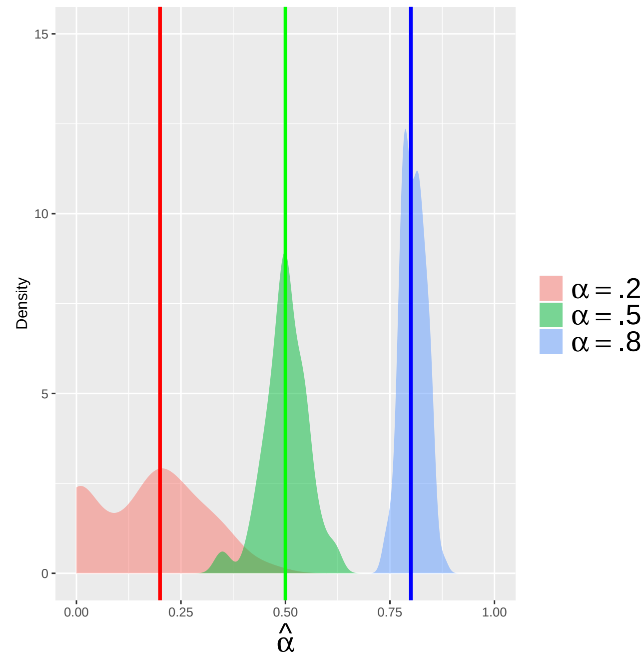
Figure 1a shows how the precision of parameter estimates changes as a function of the generating value of α. We randomly selected 200 stimuli from the set described in the previous section, and simulated choices from 300 CPT DMs on each of these stimuli. 100 DMs had a true α value of .2, 100 had a true α of .5, and 100 had a true α of .8. Notably, the estimates are most tightly clustered around their true value when $\alpha = .8$. In contrast, when $\alpha = .2$, the estimates are much more variable.

Visualizing the recoverability of a parameter value

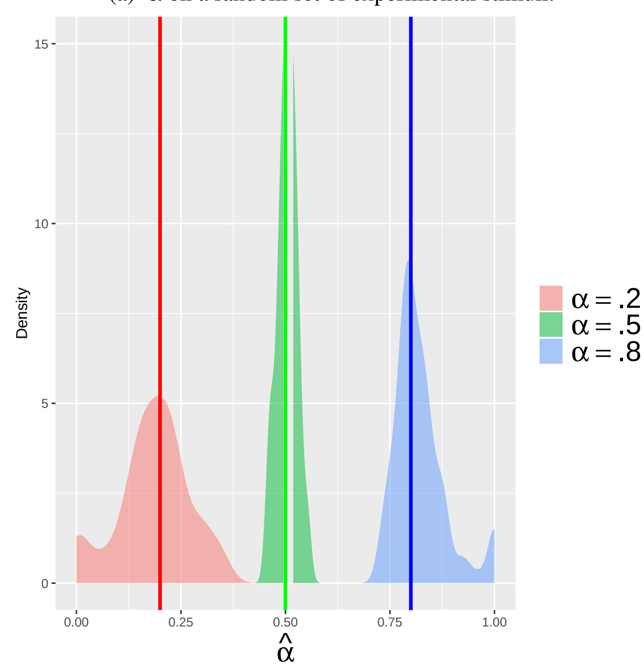
Why do we observe this asymmetry between the recoverability of $\alpha = .2$ and $\alpha = .8$? The previous sections highlights that parameter estimation routines rely on observed behavior to discriminate between parameter values. When behavior changes more rapidly in certain parts of the parameter space than in others, different parameter values will lead to more or less precise estimation.

Figure 2 shows how $P(G_{\cdot,1})$, the probability that a DM will select G_1, changes as a function of the DM's α value. Recall that the probability of making any choice is a function of the difference $V(G_{\cdot,1}) - V(G_{\cdot,2})$. Here, we've restricted $\gamma = .5$ and $\varepsilon = .5$. Each red line corresponds to a stimulus i. On the x-axis are the values we allow α to take, and on the y-axis is the difference $V(G_{i,1}) - V(G_{i,2})$. The lines are trajectories that illustrate how the differences between $V(G_{\cdot,1})$ and $V(G_{\cdot,2})$ become amplified as the value of α changes. The black lines delineate *choice probability contours*: the values of $P(G_{\cdot,1})$ that correspond to the values on the y-axis.

For the hypothetical meteorologist who wants to infer temperature, discriminating between 30°F and 40°F is straightforward because of the dramatic changes in precipitation patterns in this interval. When estimating CPT parameters, ideally changes in the parameter values will correspond to steep



(a) $\hat{\alpha}$ on a random set of experimental stimuli.



(b) $\hat{\alpha}$ on a set of experimental stimuli curated to maximize the recoverability of $.1 \leq \alpha \leq .3$.

Figure 1: Densities of the distribution of estimates of α, $\hat{\alpha}$, in a one-parameter version of CPT when the true generating value of α is .2 (red), .5 (green) and .8 (blue).

changes in choice probabilities. This framework allows us to identify ranges of α that are easier to recover by identifying where segments of trajectories cut most dramatically across choice probability contours.

Notice that when α is small ($\alpha < .5$) most of the movement

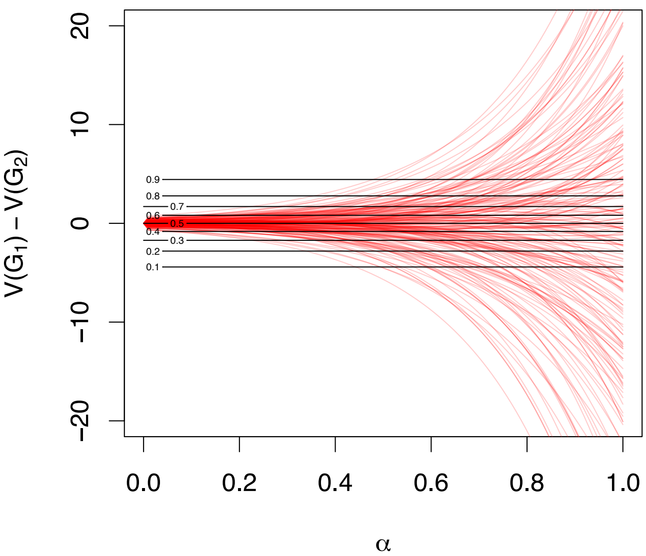


Figure 2: Visualizing the recoverability of α. Each red line corresponds to a randomly-selected stimulus i, and shows how the difference $V(G_{i,1}) - V(G_{i,2})$ changes as α is increased ($\gamma = \varepsilon = .5$). The contour lines denote the values of $P(G_{\cdot,1})$ associated with each value on the y-axis.

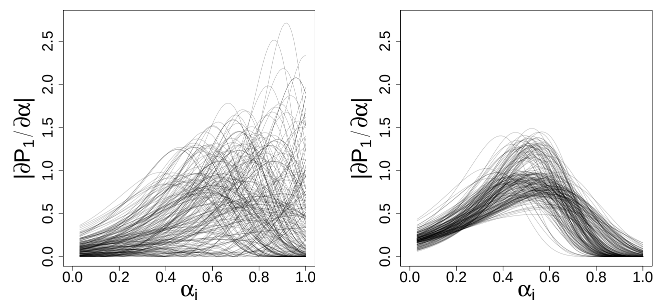
is *between* contours: Incremental changes in parameter values barely shift choice probabilities. Compare this to the dramatic movement when α is high ($\alpha > .5$). For this range of αs, the trajectories cut aggressively across contours, indicating that incremental changes of α in this range correspond to large changes in choice probabilities.[2]

We expect that stimuli that cross more contours yield more information about the value of α. Figure 3a illustrates the formalization of the number of contours a stimulus crosses. It plots the absolute value of the finite differential $\frac{\partial P(G_{\cdot,1})}{\partial \alpha}$ across the range of plausible values of α. Each line corresponds to one of the 200 randomly-selected stimuli. The finite differential is the difference between $P(G_{\cdot,1})$ calculated at $\tilde{\alpha}$ and $P(G_{\cdot,1})$ calculated at $\tilde{\alpha} - \delta$, where δ is a very small increment (and γ and ε are held constant). In general, we can calculate the differential of $P(G_{\cdot,1})$ with respect to any parameter. We hereafter simplify terminology and write $\partial P_1 / \partial \theta$ where θ is the parameter the differential is calculated with respect to.

In other words, the y-axis shows the number of contours each stimulus crosses when α increases by some small amount. We expect that a value of α for which the stimuli cut across many contours would be more precisely estimated than a value of α for which the stimuli did not move very much. We therefore favor regions where the absolute value of the differential is far from 0.

For low values of α, the lines cluster around 0, but tend to diverge for higher values of α. These higher absolute values

[2]We speculate that because $v(X_\cdot)$ grows monotonically with α, large values of α magnify the difference $V_{\cdot,1} - V_{\cdot,2}$. The rate at which this difference grows accelerates as α increases, cutting across more contours.



(a) A random set of experimental stimuli.

(b) Stimuli with the highest $|\partial P_1 / \partial \alpha|$ for $.1 \leq \alpha \leq .3$.

Figure 3: $|\partial P_1 / \partial \alpha|$. Each line corresponds to a single experimental stimulus.

of $\partial P_1 / \partial \alpha$ are consistent with the more precise estimates of higher absolute values of α we display in Figure 1a.[3]

Effect of adjusting the experiment

$\partial P_1 / \partial \alpha$ is a measure of the "suddenness" of change in predicted behavior, and can be thought of as a crude estimate of the parameter discrimination of a stimulus or set of stimuli at a particular value of α. We expected that if we adjusted the stimulus set so the aggregate $\partial P_1 / \partial \alpha$ was more concentrated around lower values of α, this would mitigate the structural asymmetry between estimating high and low values of α.

For each stimulus, we calculated the absolute value of $\partial P_1 / \partial \alpha$ at each value of α. To select observations that would be especially helpful in discriminating between low values of α, we created an aggregated differential for each stimulus by stacking the values of $|\partial P_1 / \partial \alpha|$ such that $.1 \leq \alpha \leq .3$, and selected the 200 stimuli with the largest aggregated differential.

Figure 3b shows the values of $|\partial P_1 / \partial \alpha|$ for these 200 stimuli. Interestingly, there appear to be no stimuli where $|\partial P_1 / \partial \alpha|$ peaks at $\alpha = .2$. Instead, we isolated stimuli where the absolute value of the differential peaks around $\alpha = .5$, which tends to "pull" some of the mass at $\alpha = .2$ away from 0.

Figure 1b shows the distribution of parameter estimates when choices are simulated using this set of curated stimuli. As we predicted, the estimates when $\alpha = .2$ and $\alpha = .5$ seem to have a lower variance than when the stimuli were selected randomly, particularly estimates when $\alpha = .5$. However, the estimates when $\alpha = .8$ are slightly more dispersed.

This is consistent with the pattern in Figure 3b: the values of $\partial P_1 / \partial \alpha$ are, in aggregate, further from 0 when $\alpha \approx .5$, and are relatively flatter when $\alpha \approx .8$. This suggests that this measurement can elucidate properties of an experimental design that map onto empirical asymmetries in the recoverability of parameter values.

[3]While these results are calculated on the basis of randomly-selected gambles in the domain of gains, we see the same general trend in the finite differentials of previously-published data sets, such as those analyzed by Erev et al. (2002), Stott (2006) and Glöckner and Pachur (2012).

Two-parameter model

In the previous section, we showed how the methods we propose can anticipate structural asymmetries in the recoverability of different values of α. However, the parameters of CPT are interdependent (Li, Lewandowsky, & DeBrunner, 1996; Scheibehenne & Pachur, 2015), and we expect the dynamics of the individual parameters to interact. In this section, we extend our analysis to a two-parameter version of CPT. We again fix ε at .5, and allow both α and γ to vary.

Figure 4 shows the distribution of estimates of γ when $\alpha =$ .5 and when $\alpha = .8$. (Unlike in the analysis above, here α and γ are estimated jointly.) Notice the interaction between the values of α and γ: Estimates of γ are more precise when α is high, and especially so when α is high and γ is low.

The same interaction can be seen from the corresponding $|\partial P_1/\partial\gamma|$s. Figure 3 shows plots of the finite differentials with respect to one parameter. Figure 5a shows the two-dimensional analog: a surface of the average value of $|\partial P_1/\partial\gamma|$. The interaction we observed in Figure 4 is also noticeable here: The surface is higher for higher values of α, and especially so when α is high and γ is low.

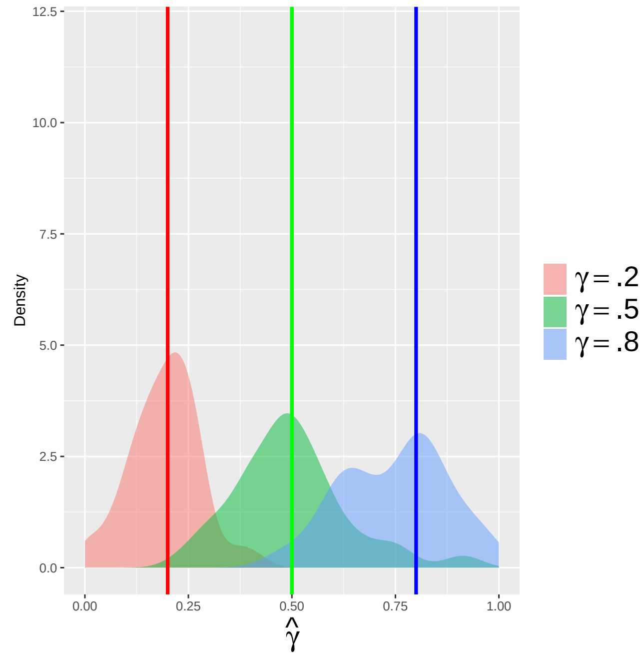
Using a procedure similar to the one we followed for the one-parameter model, we attempted to mitigate this asymmetry in the estimation of γ. We selected the stimuli with the highest $|\partial P_1/\partial\gamma|$ when α was constrained to equal .5. The parameter estimation results using this curated design are displayed in Figure 6. We were able to enhance the recoverability of γ when the true value of γ is .2, but do not otherwise observe a substantial increase in the precision of the estimates.

Figure 5b provides insight as to why we see this pattern. The $|\partial P_1/\partial\gamma|$ surface climbs most dramatically when γ is low, translating into the relatively more precise estimates shown in Figure 6. Again, visualizing the choice probability surface—the relationship between changes in parameter values and changes in behavior—facilitated diagnosis, interpretation and anticipation of asymmetries in parameter estimation.

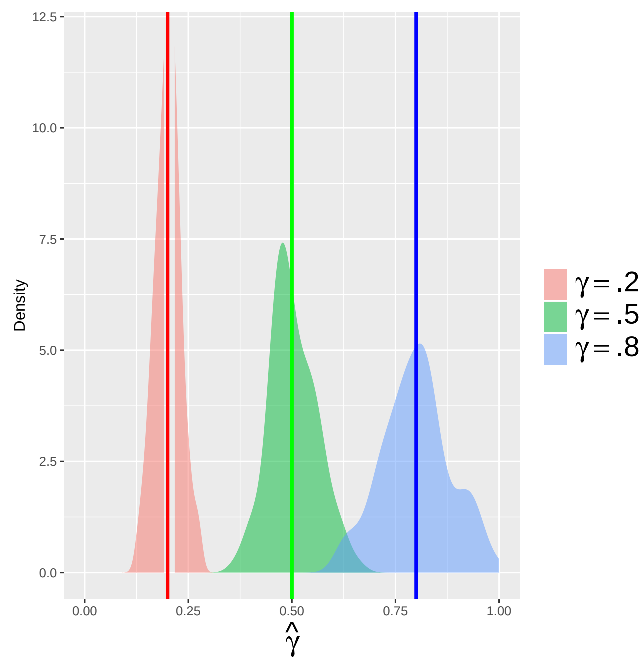
Discussion

We presented methods to visualize the interplay between an experimental design and a model's parameters, and to facilitate understanding of model-stimulus relationships.

While we focused discussion on CPT, the intuitions we established for structural asymmetries in parameter estimation towards the beginning of the paper would apply more broadly to non-linear cognitive models, especially to models of decision-making, models of categorization and other models that are often calibrated using discrete observations. In general, consideration of and transparency about the diagnosticity of an experimental design is an important part of robust modeling practices (Broomell et al., 2019). While we applied these methods to a one- and two-parameter model, advances in high-dimensional data visualization will likely make scaling the approach to more complex models increasingly easier (Blaha, 2019; Glendenning, Wischgoll, Harris, Vickery, & Blaha, 2016).



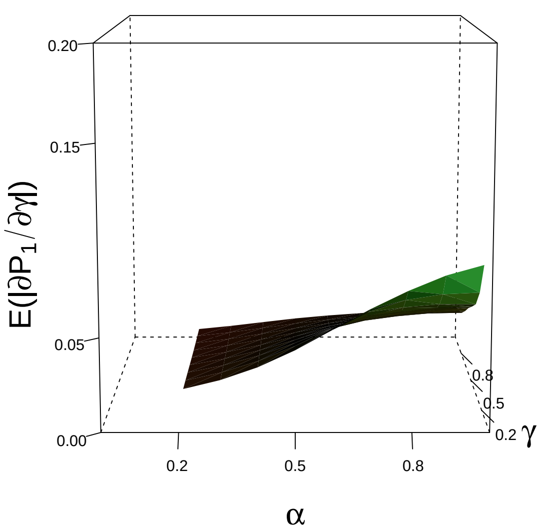
(a) $\alpha = .5$.



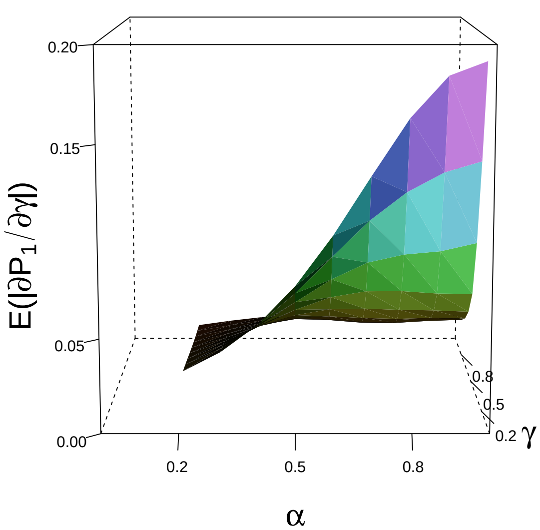
(b) $\alpha = .8$.

Figure 4: Densities of the distribution of estimates of γ, $\hat{\gamma}$, in a two-parameter version of CPT when the true generating value of γ is .2 (red), .5 (green) and .8 (blue). The experimental stimuli are randomly selected. The generating value of α is either .5 (top) or .8 (bottom).

We are far from the first to point out challenges in the identification of CPT's parameters (although we have not come across another reference to the particular structural asymmetries we discuss). Researchers have advocated for the



(a) Mean $|\partial P_1/\partial\gamma|$ of a random set of experimental stimuli.



(b) Mean $|\partial P_1/\partial\gamma|$ of a set of experimental stimuli curated to maximize the recoverability of γ conditional on $\alpha = .5$.

Figure 5: Mean, across stimuli, of the absolute value of the $\partial P_1/\partial\gamma$.

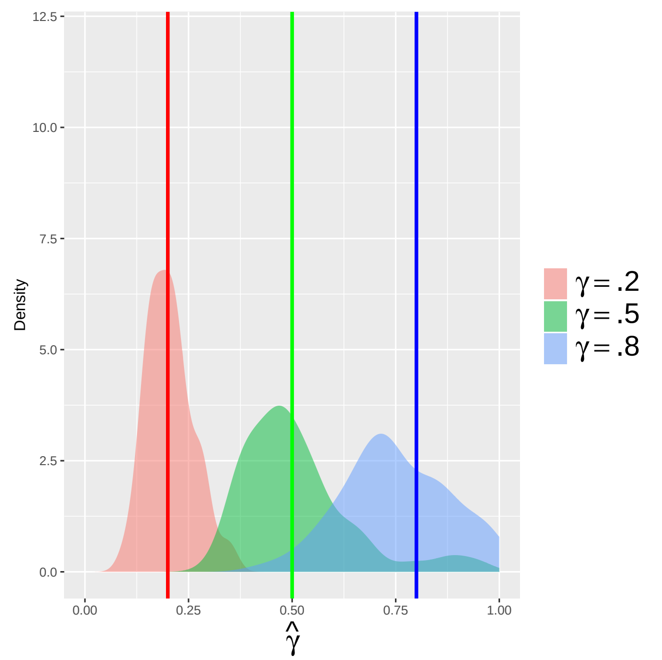


Figure 6: Densities of the distribution of estimates of γ, $\hat{\gamma}$, in a two-parameter version of CPT when the true generating value of γ is .2 (red), .5 (green) and .8 (blue), and the generating value of $\alpha = .5$. The experimental stimuli are curated to maximize the recoverability of γ when $\alpha = .5$.

use of hierarchical Bayesian approaches to the estimation of CPT parameters in favor of maximum likelihood methods (Jarnebrant, Toubia, & Johnson, 2009; Nilsson, Rieskamp, & Wagenmakers, 2011; Scheibehenne & Pachur, 2015; Toubia et al., 2013). These studies highlight the ability of hierarchical Bayesian methods to introduce shrinkage into parameter estimates, reducing the influence of outlying estimates. In addition, hierarchical Bayesian methods make explicit the uncertainty in and relationships between parameter estimates (Scheibehenne & Pachur, 2015; Wetzels, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2010). We see the hierarchical Bayesian approach as an additional, important tool that can both facilitate parameter identification and help researchers visualize a form of uncertainty in their estimates. While we used maximum likelihood for this demonstration, extensions of this work could use other estimation procedures, like hierarchical Bayesian techniques, to generate the distribution of parameter estimates shown in Figures 1, 4 and 6. The method of estimation and the experimental design are each one of many components of parameter recovery studies, and both importantly contribute to the conclusions drawn by researchers.

We hope our work both promotes a more nuanced understanding of the measurement process of parameter estimation, and provides a step forward in developing ways to evaluate experimental designs in a way that accommodates these nuances.

References

Blaha, L. M. (2019, December). We Have Not Looked at Our Results Until We Have Displayed Them Effectively: A Comment on Robust Modeling in Cognitive Science. *Computational Brain & Behavior*, *2*(3-4), 247–250. doi: 10.1007/s42113-019-00059-6

Broomell, S. B., & Bhatia, S. (2014). Parameter Recovery for Decision Modeling Using Choice Data. *Decision*, *1*(4), 252–274.

Broomell, S. B., Sloman, S. J., Blaha, L. M., & Chelen, J. (2019, December). Interpreting Model Comparison Requires Understanding Model-Stimulus Relationships. *Computational Brain & Behavior*, *2*(3-4), 233–238. doi: 10.1007/s42113-019-00052-z

Cavagnaro, D. R., Myung, J. I., Pitt, M. A., & Kujala, J. V. (2010, April). Adaptive Design Optimization: A Mutual Information-Based Approach to Model Discrimination in Cognitive Science. *Neural Computation*, *22*(4), 887–905. doi: 10.1162/neco.2009.02-09-959

Erev, I., Roth, A. E., Slonim, R., & Barron, G. (2002). Combining a Theoretical Prediction with Experimental Evidence. *SSRN Electronic Journal*. doi: 10.2139/ssrn.1111712

Glendenning, K., Wischgoll, T., Harris, J., Vickery, R., & Blaha, L. (2016, February). Parameter Space Visualization for Large-scale Datasets Using Parallel Coordinate Plots. *Electronic Imaging*, *2016*(1), 1–8. doi: 10.2352/ISSN.2470-1173.2016.1.VDA-490

Glöckner, A., & Pachur, T. (2012, April). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, *123*(1), 21–32. doi: 10.1016/j.cognition.2011.12.002

Jarnebrant, P., Toubia, O., & Johnson, E. (2009, November). The Silver Lining Effect: Formal Analysis and Experiments. *Management Science*, *55*(11), 1832–1841. doi: 10.1287/mnsc.1090.1076

Kim, W., Pitt, M. A., Lu, Z.-L., Steyvers, M., & Myung, J. I. (2014, November). A Hierarchical Adaptive Approach to Optimal Experimental Design. *Neural Computation*, *26*(11), 2465–2492. doi: 10.1162/NECO_a_00654

Li, S.-C., Lewandowsky, S., & DeBrunner, V. E. (1996). Using parameter sensitivity and interdependence to predict model scope and falsifiability. *Journal of Experimental Psychology: General*, *125*(4), 360–369. doi: 10.1037/0096-3445.125.4.360

Myung, J. I., & Pitt, M. A. (2009). Optimal experimental design for model discrimination. *Psychological Review*, *116*(3), 499–518. doi: 10.1037/a0016104

Navarro, D. J., Pitt, M. A., & Myung, I. J. (2004, August). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology*, *49*(1), 47–84. doi: 10.1016/j.cogpsych.2003.11.001

Nilsson, H. k., Rieskamp, J., & Wagenmakers, E.-J. (2011, February). Hierarchical Bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, *55*(1), 84–93. doi: 10.1016/j.jmp.2010.08.006

Prelec, D. (1998, May). The Probability Weighting Function. *Econometrica*, *66*(3), 497. doi: 10.2307/2998573

Scheibehenne, B., & Pachur, T. (2015, April). Using Bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychonomic Bulletin & Review*, *22*(2), 391–407. doi: 10.3758/s13423-014-0684-4

Stewart, N., Canic, E., & Mullett, T. L. (2019, November). *On the futility of estimating utility functions: Why the parameters we measure are wrong, and why they do not generalize* (Preprint). PsyArXiv. doi: 10.31234/osf.io/qt69m

Stott, H. P. (2006, March). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty*, *32*(2), 101–130. doi: 10.1007/s11166-006-8289-6

Toubia, O., Johnson, E., Evgeniou, T., & Delquié, P. (2013, March). Dynamic Experiments for Estimating Preferences: An Adaptive Method of Eliciting Time and Risk Parameters. *Management Science*, *59*(3), 613–640. doi: 10.1287/mnsc.1120.1570

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, *5*, 27.

Wetzels, R., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2010, February). Bayesian parameter estimation in the Expectancy Valence model of the Iowa gambling task. *Journal of Mathematical Psychology*, *54*(1), 14–27. doi: 10.1016/j.jmp.2008.12.001