Fast and Simple Adaptive Elicitations

Nicolò Bertani *[1], Enrico Diecidue[2], Patrice Perny[3], and Paolo Viappiani[4]

[1]Católica Lisbon School of Business & Economics, Lisbon, Portugal
[2]Decision Sciences, INSEAD, Fontainebleau, France
[3]LIP6, Sorbonne University, CNRS, Paris, France
[4]LAMSADE, CNRS, Université Paris-Dauphine PSL, Paris, France

For Online Publication

Online Appendix

OA.A Additional illustrations of the elicitation of the probability weighting function

OA.A.1 Additional cases from Abdellaoui (2000, Table 9)

Figures OA.1figure.1, OA.2figure.2, and OA.3figure.3 complement Figure 6 in §4.1 and illustrate the step-by-step elicitation of the remaining estimated functional forms of Abdellaoui (2000, Table 9).

The output associated with each one of these simulated recoveries is illustrated in Figure OA.4figure.4, OA.5figure.5, and OA.6figure.6, respectively. These figures are analogous to Figure 7 again in §4.1.

*Corresponding author: nbertani@ucp.pt

Figure OA.1: Step-by-step elicitation - GE with $r = .84, s = .65$.

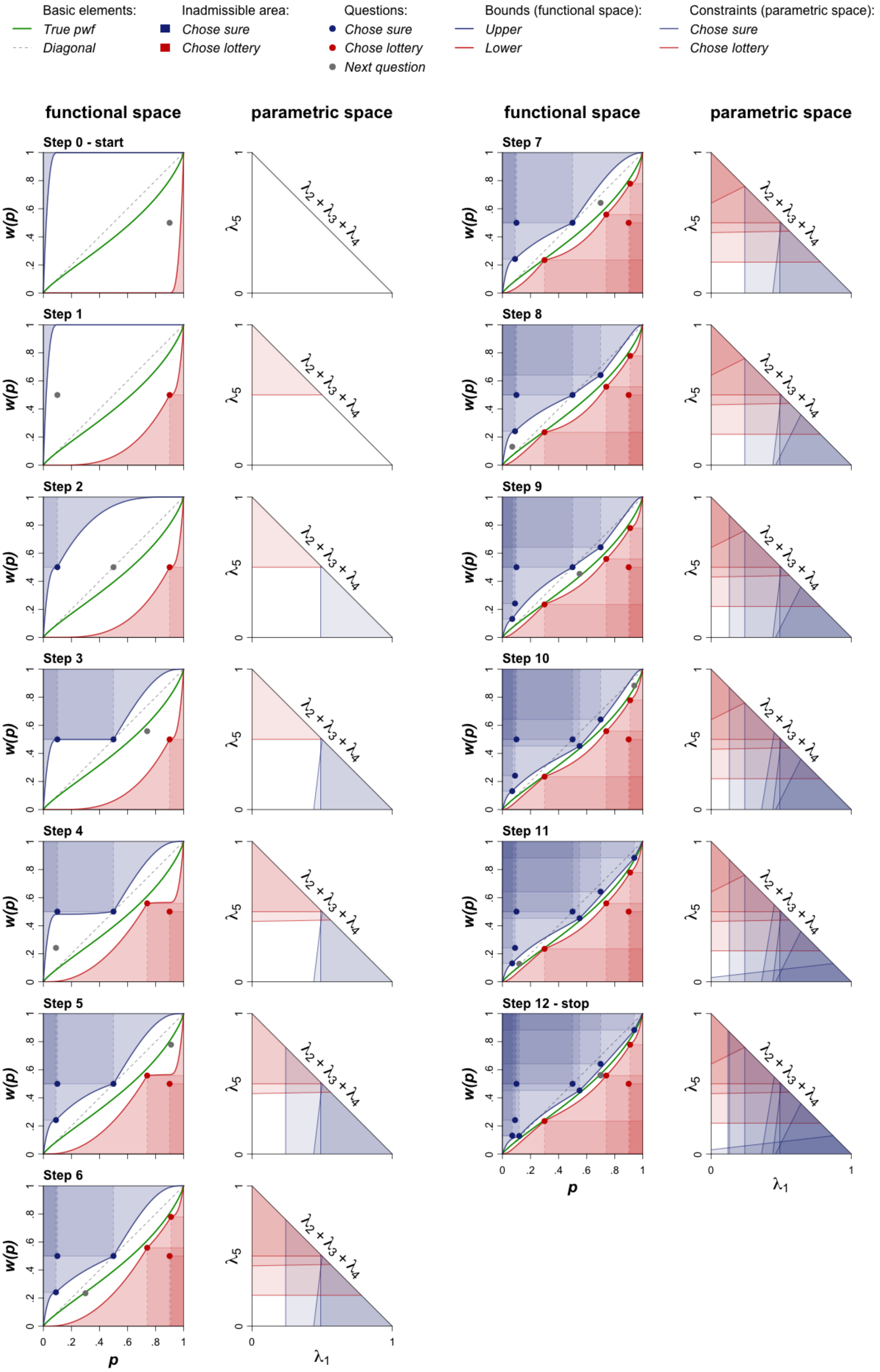


Figure OA.2: Step-by-step elicitation - TK with $r = .6$.

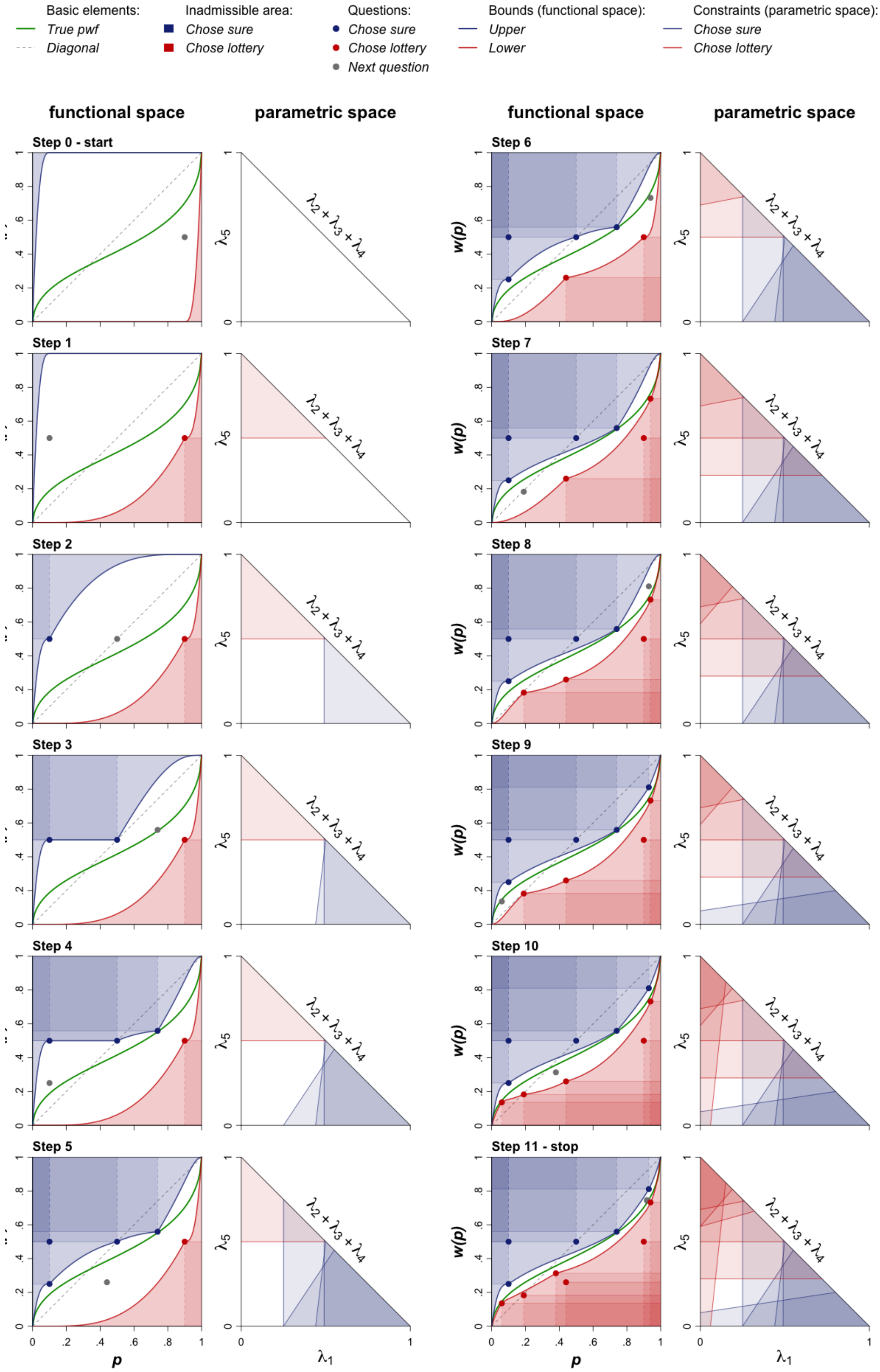


Figure OA.3: Step-by-step elicitation - TK with $r = .7$.

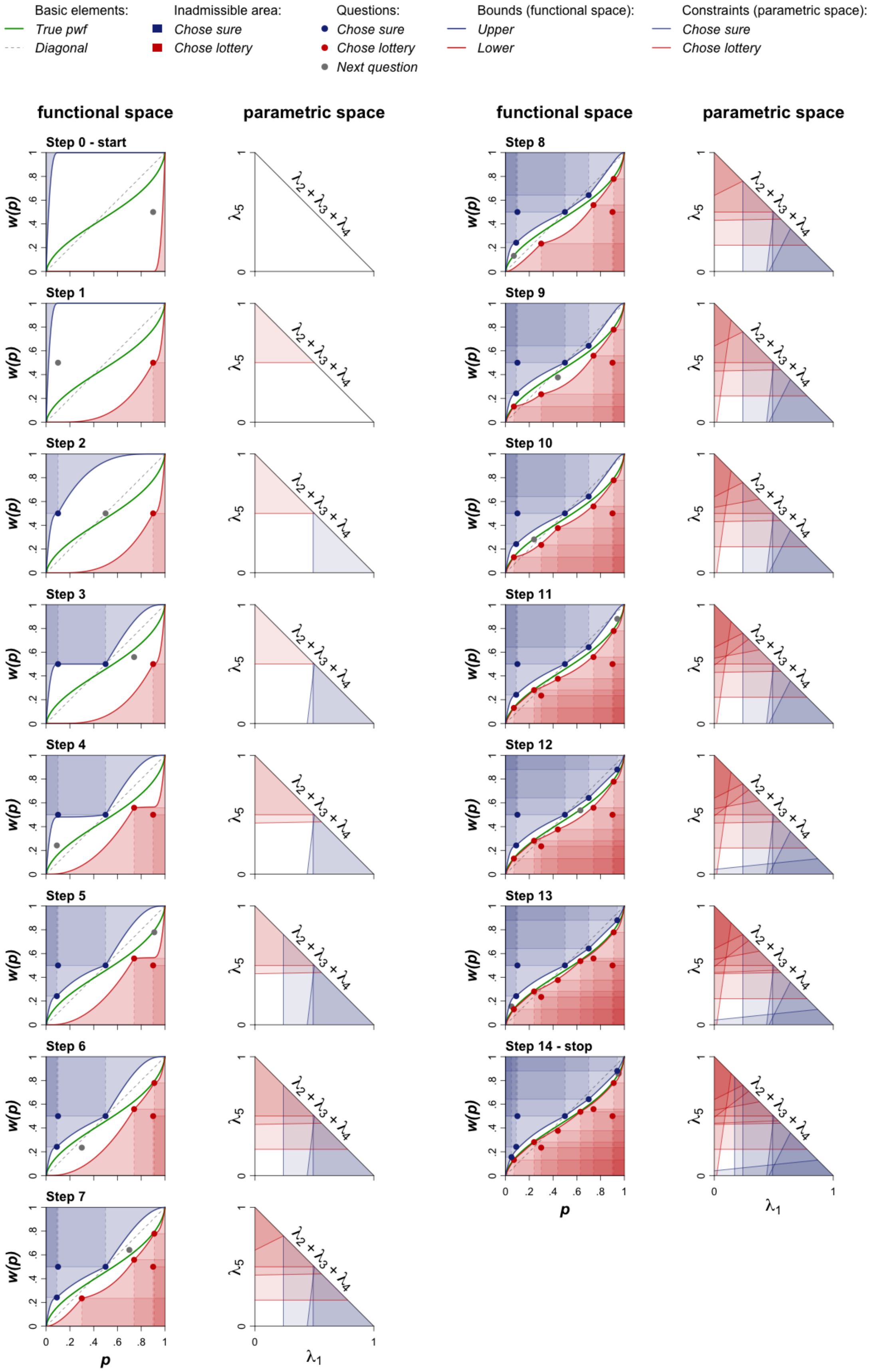


Figure OA.4: Final output - GE with $r = .84, s = .65$.

(a) functional space. (b) parametric space. (c) distances at termination.

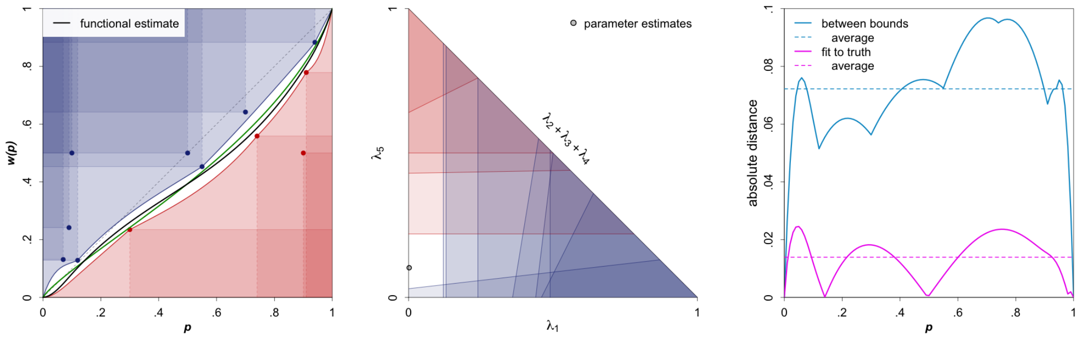


Figure OA.5: Final output - TK with $r = .6$.

(a) functional space. (b) parametric space. (c) distances at termination.

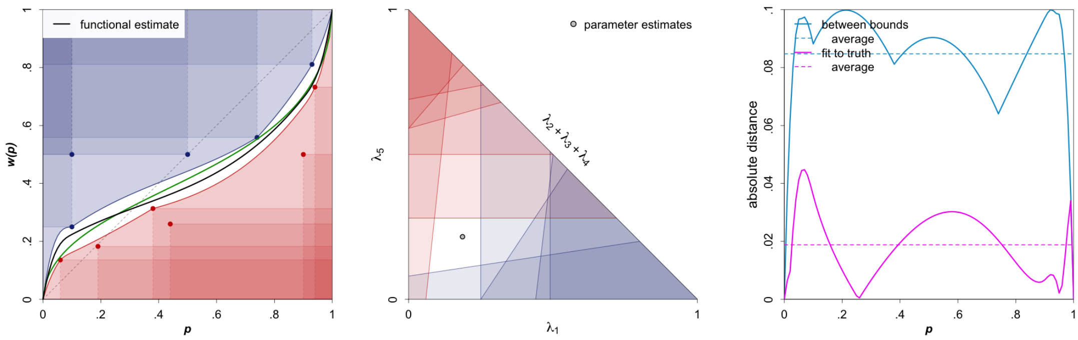
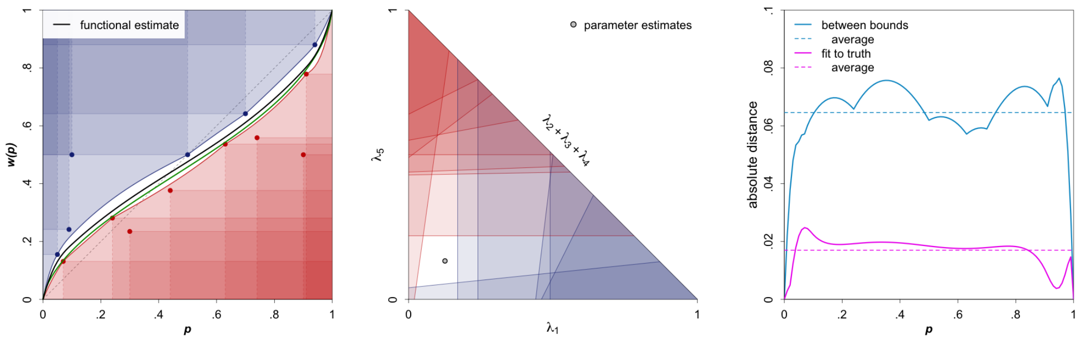


Figure OA.6: Final output - TK with $r = .7$.

(a) functional space. (b) parametric space. (c) distances at termination.



OA.A.2 Elicitation with a coarser discretization

In both the simulated and experimental elicitation of the probability weighting function, §4, we use a discretization of probability to percentage points, i.e. $\mathcal{D} = \{0, .01, .02, \ldots, .99, 1\}$. For the simulated elicitation of the utility function, §OA.D.2subsection.4.2, the sure amount is discretized to 5 cent intervals, i.e. $\mathcal{D} = \{.1, .15, \ldots, 3.8, 3.85\}$. Differences of one percentage point or 5 cents might be difficult to perceive and understand by a human respondent, thus it is natural to wonder how FSE would work under a coarser discretization. To address this question, we again recover the estimated probability weighting functions of Abdellaoui (2000, Table 9), this time using a discretization to probability deciles, i.e. $\mathcal{D} = \{0, .1, .2, \ldots, .9, 1\}$. This is the same discretization assumed, for instance, in the adaptive elicitation of Cavagnaro et al. (2013a,b).

The step-by-step elicitation is illustrated in Figure OA.7figure.7, OA.8figure.8, OA.9figure.9, and OA.10figure.10, with the associated final output in Figure OA.11figure.11, OA.12figure.12, OA.13figure.13, and OA.14figure.14, respectively. Note that under this coarser discretization, the elicitation requires a larger number of questions, since the sequencing is constrained to a smaller number of possible values. Nonetheless, as it can be noticed in subfigures OA.11csubfigure.11.3, OA.12csubfigure.12.3, OA.13csubfigure.13.3, and OA.14csubfigure.14.3, the maximal distance between the bounds is still decreased to values smaller than the threshold ($\epsilon = .1$) over the entire support. Similarly, the accuracy of the elicitation is not negatively impacted: in fact, due to the larger number of questions being asked, the estimated shapes are even closer to the true underlying function than in subfigures 7c, OA.4csubfigure.4.3, OA.5csubfigure.5.3, and OA.6csubfigure.6.3, respectively.

Figure OA.7: Elicitation with coarser discretization - GE with $r = .65, s = .6$.

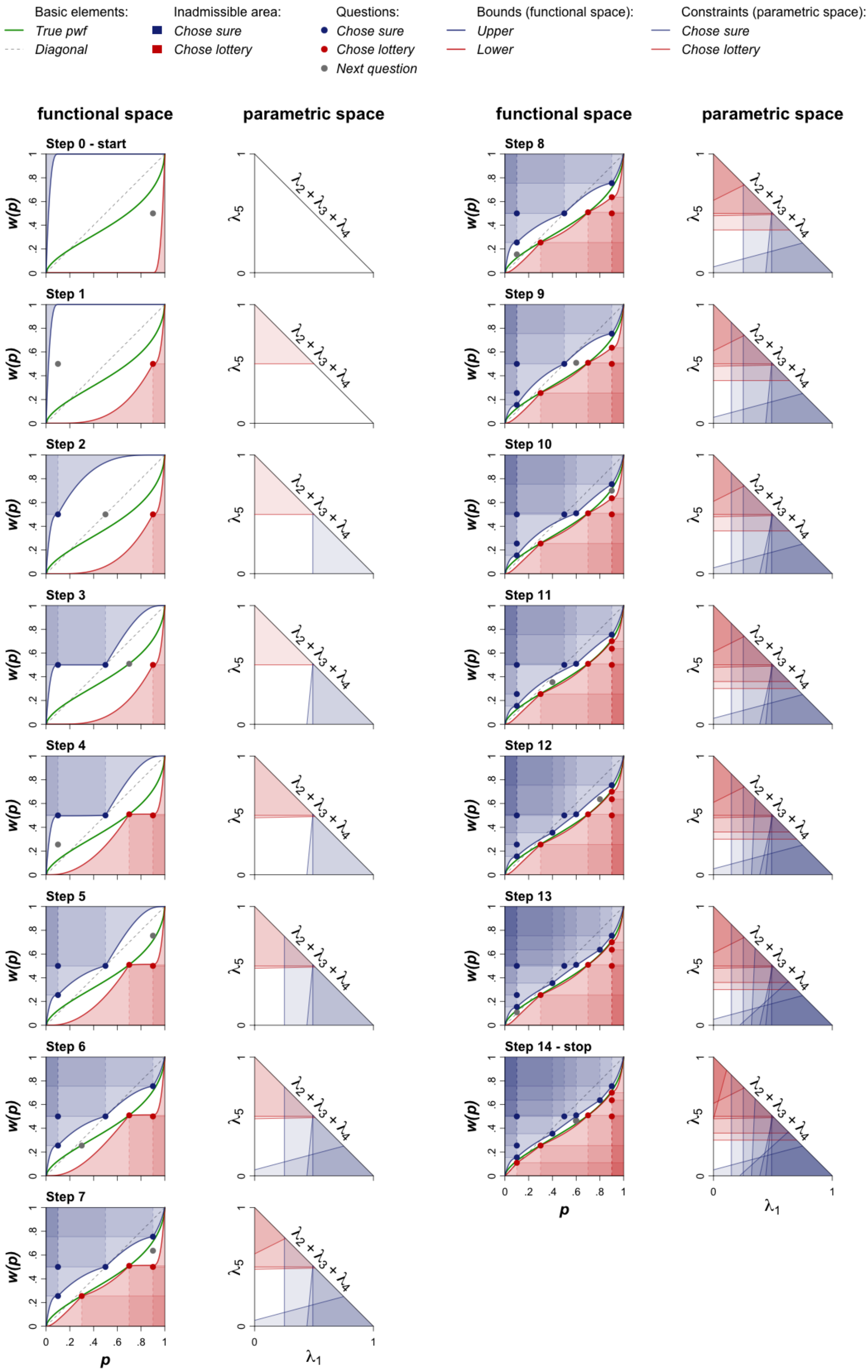


Figure OA.8: Elicitation with coarser discretization - GE with $r = .84$, $s = .65$.

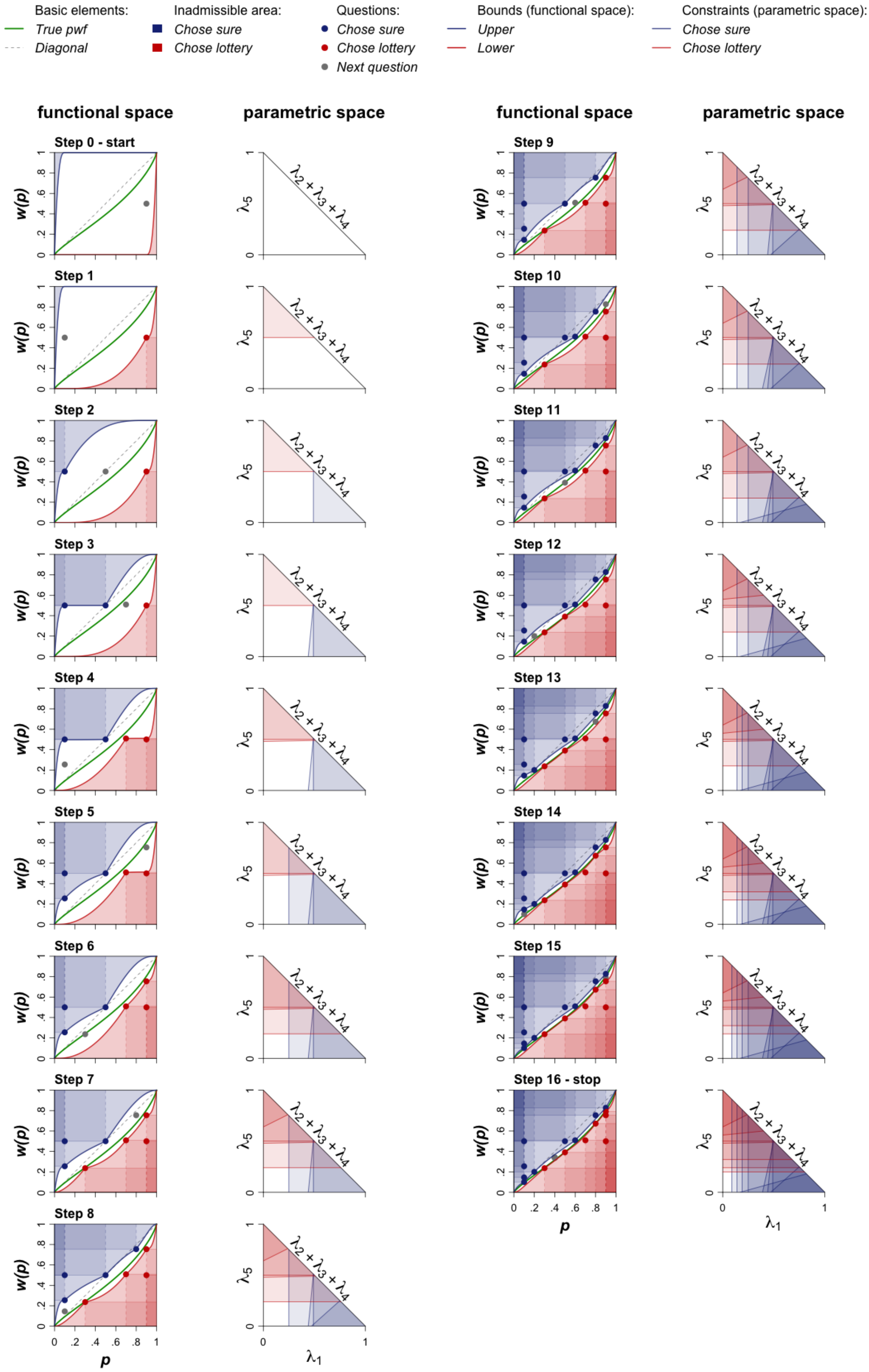


Figure OA.9: Elicitation with coarser discretization - TK with $r = .6$.

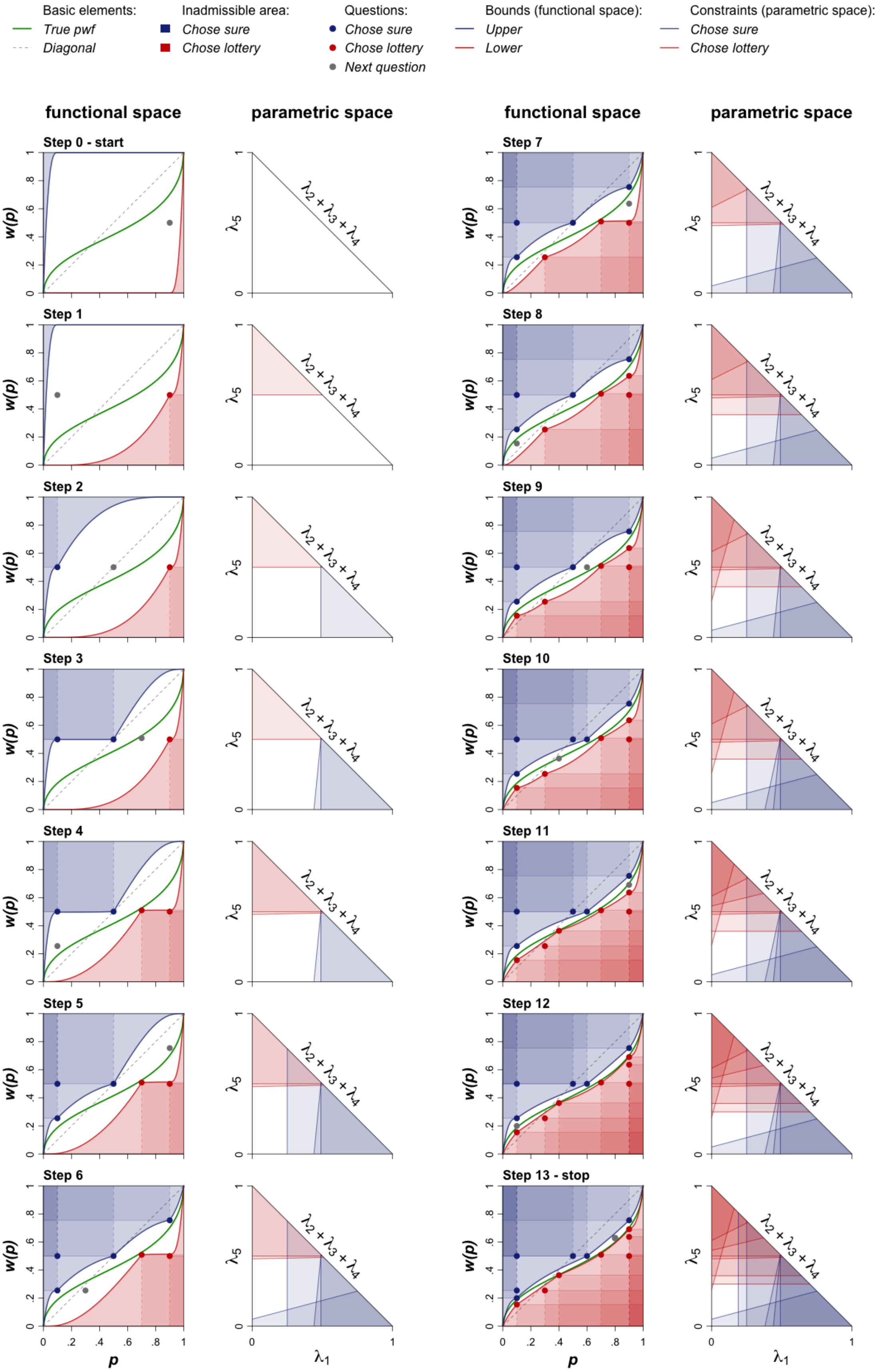


Figure OA.10: Elicitation with coarser discretization - TK with $r = .7$.

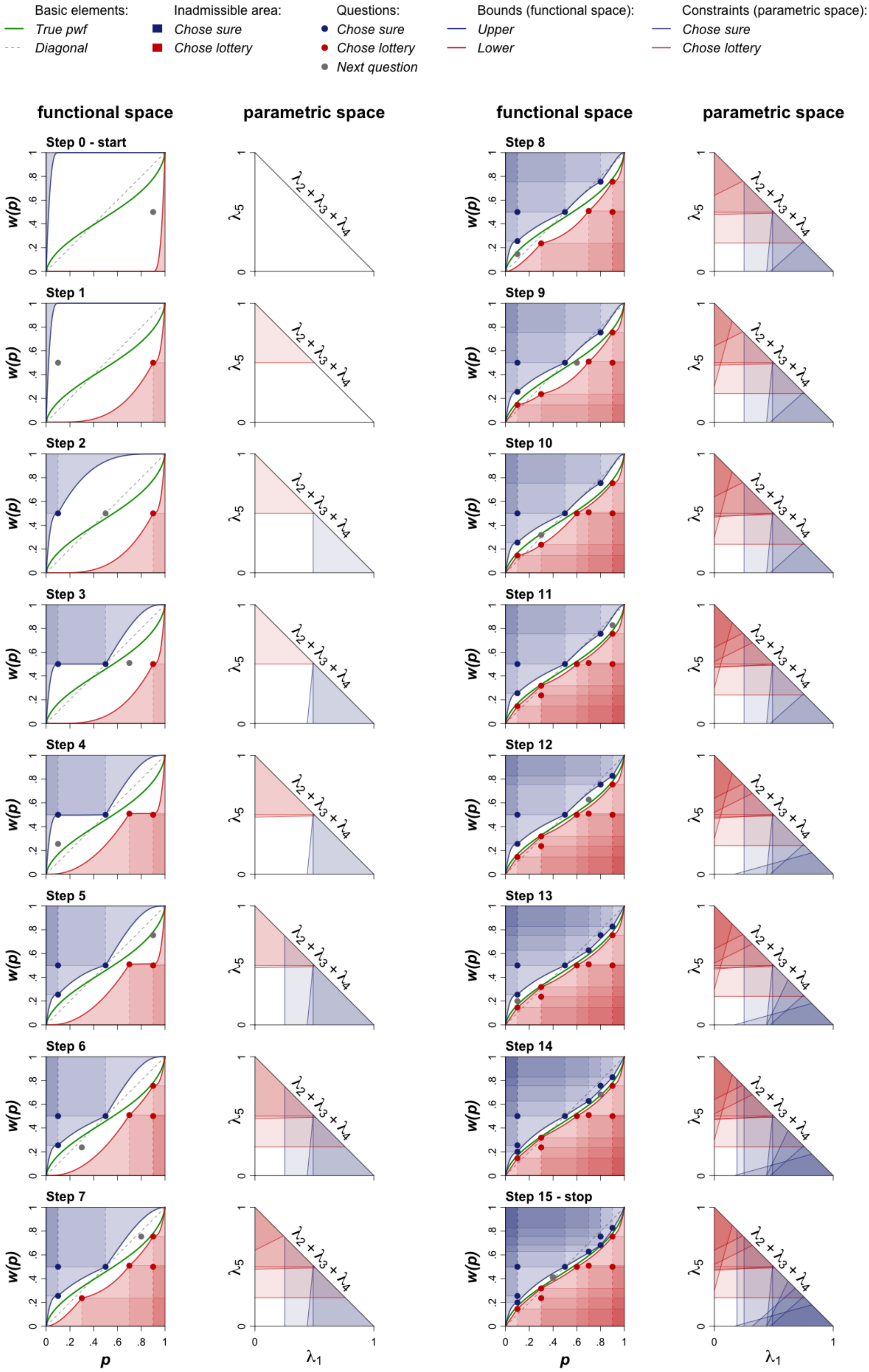


Figure OA.11: Final output with coarser discretization - GE with $r = .65, s = .6$.

(a) functional space. (b) parametric space. (c) distances at termination.

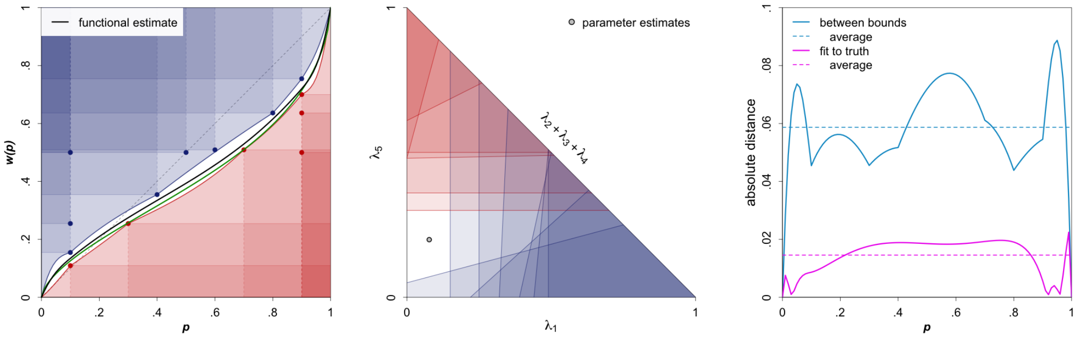


Figure OA.12: Final output with coarser discretization - GE with $r = .84, s = .65$.

(a) functional space. (b) parametric space. (c) distances at termination.

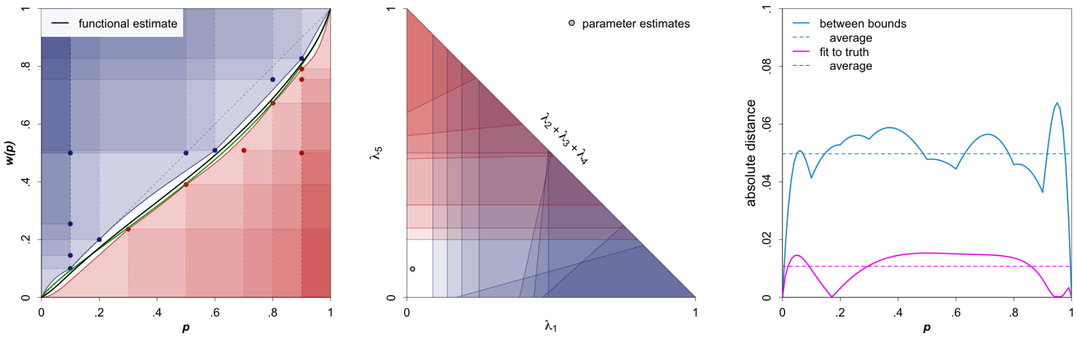


Figure OA.13: Final output with coarser discretization - TK with $r = .6$.

(a) functional space. (b) parametric space. (c) distances at termination.

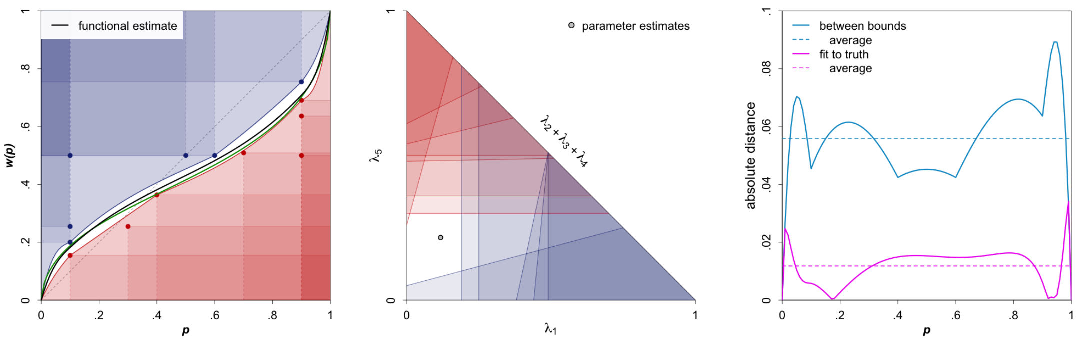
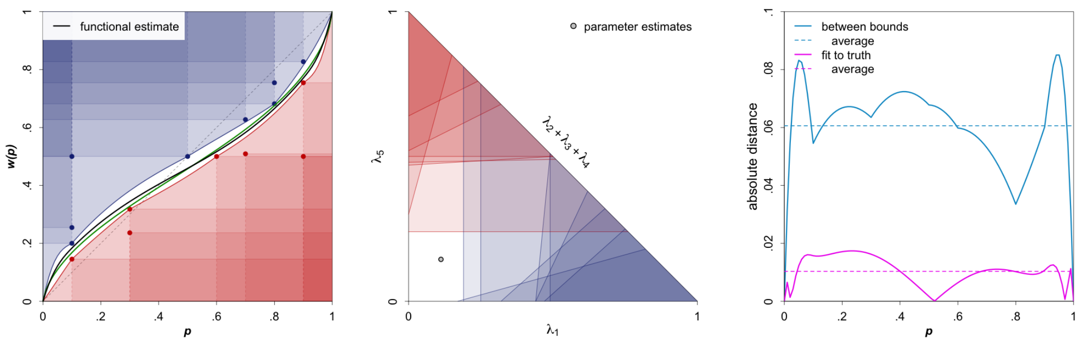


Figure OA.14: Final output with coarser discretization - TK with $r = .7$.

(a) functional space. (b) parametric space. (c) distances at termination.



OA.B Complement to the laboratory experiment

OA.B.1 Details of experiment

We tested the empirical validity of our procedure with 71 incentivized subjects (40 women, 31 men, average age: 22.6). The study was approved by the INSEAD Institutional Review Board and the data was collection at the INSEAD-Sorbonne Université Behavioural Lab in Paris. The subjects received a fixed participation fee of € 5 plus additional payments depending on a randomized incentive scheme, as per standard practice (Starmer and Sugden, 1991, Cubitt et al., 1998). The random compensation amounted to € 857 (eight participants gained an average of € 107 extra). The subjects underwent the procedure autonomously in group sessions. The instructions, which are available in Online Appendix OA.B.2subsection.2.2, were limited to a slim 10-slide deck comprising fewer than 300 words and 5 illustrative choice questions. A lab technician was present during the procedure for software support. We set the procedure's flexible components as follows. The fixed lottery outcomes x and y were set to € 120 and € 10, respectively. The rationale behind this choice was to strike a sensible balance between sensitivity to the outcomes and acceptability of the linear utility approximation (Birnbaum 2008 recommends $\$1 < y \leq x < \$150$). Consistently, we intentionally excluded the € 0 outcome (Lopes and Oden, 1999, Bateman et al., 2007, Diecidue et al., 2015). The subdivision was defined as $\boldsymbol{\xi} = \{.1, .9\}$, determining a 5 I-spline basis. Given this choice, we have local and independent descriptive power for possibility up to .1 and a certainty effect from .9, with more smoothness in $(.1, .9)$, to adequately capture insensitivity. Additionally, this subdivision contributes to limiting the number of questions. We chose $\epsilon = .1$, which meant that the bounds over the whole support would be ubiquitously tighter than the windows for indifference points that a one-stage PL would define at only N points in the support (as illustrated in Figures 6 and 2b). To check response consistency, the third choice question presented to every subject was repeated as a concluding question.

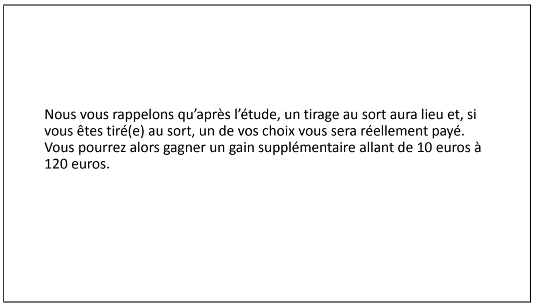
OA.B.2 Experimental instructions and interface

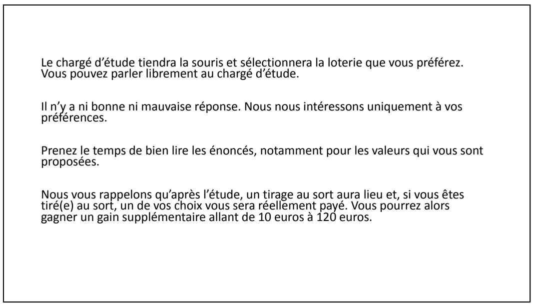
Our instructions consisted in a 10-slide deck comprising less than 300 words and 5 illustrative questions. They are reported in Figure OA.15figure.15. The interface of the experiment was consistent with the instructions. When displaying the options for each binary choice, the location of the sure amount and the lottery was alternated, to avoid mechanical responses. Some examples of the interface are presented in Figure OA.16figure.16.

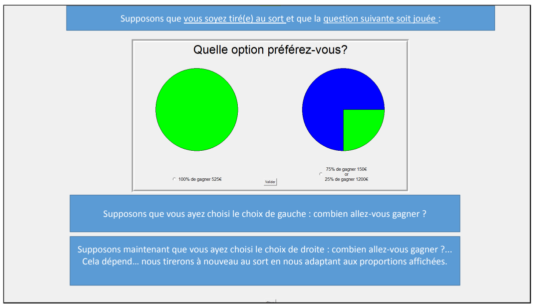
Figure OA.15: experimental instructions.

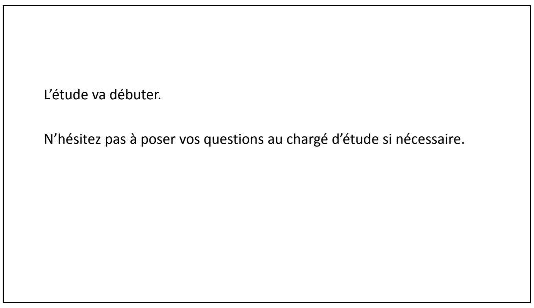
English translations.
Slide 1 (topleft): Welcome to the Choice and Lottery study. This study lasts about 10 minutes and is compensated with 4 euros. In this study, you will be asked to choose between two lotteries. There is no right or wrong answer. We are here interested in your preferences. A random draw will take place at the end of the session and, if you are selected, one of your choices will be actually paid out. You could then earn a bonus payment ranging from 10 euros to 120 euros.
Slide 2 (topright): Example question: What option do you prefer? 10% of winning €150 or 90% of winning €1200. 100% of winning €675. · If you choose the lottery on the right, how much could you win? · If you choose the lottery on the left, how much could you win?"
Slide 3 (centerleft): · Other examples:
The text in the remaining slides is analogous to that in Slide 2.

Figure OA.15: experimental instructions - continued.



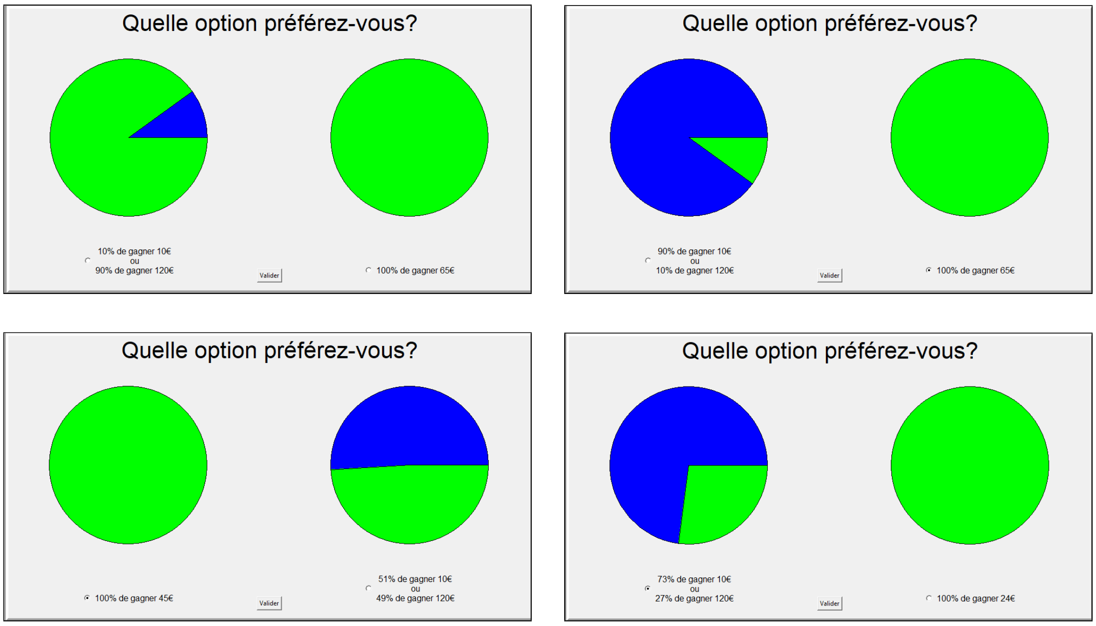






English translations.
Slide 7 (topleft): We remind you that after the study, a draw will take place, and if you are selected, one of your choices will actually be paid out to you. You could then win an additional amount ranging from 10 euros to 120 euros.
Slide 8: The study administrator will handle the mouse and select the lottery you prefer. You can speak freely with the administrator. There is no right or wrong answer. We are only interested in your preferences. Take your time to carefully read the statements, particularly the values presented to you. We remind you that after the study, a draw will take place, and if you are selected, one of your choices will actually be paid out to you. You could then win an additional amount ranging from 10 euros to 120 euros.
Slide 9 (bottomleft): Suppose you are selected in the draw and the following question is played: What option do you prefer? 100% of winning €525. 75% of winning €150 or 25% of winning €1200. Suppose you have chosen the option on the left: how much will you win? Now suppose you have chosen the option on the right: how much will you win? It depends... we will draw again, adapting to the displayed proportions.
Slide 10 (bottomright): The study is about to begin. Feel free to ask the study administrator any questions if necessary.

Figure OA.16: illustrative screenshots of the experimental interface.



English translations as in Slide 2 of Figure OA.15figure.15.

OA.C Extending FSE to handle errors

OA.C.1 *FSE with validation* and *FSE with step-back*

FSE aims to minimize the number of simple questions required to elicit a decision-theoretic function. Consequently, it assumes deterministic responses, collects no contradictory choices, and does not probe the consistency of the collected choices. As an adaptive procedure, FSE's treatment of information can be leveraged to naturally handle possible response errors. Specifically, the procedure can be extended to incorporate adaptive consistency checks and to accommodate contradictory preferences. Interestingly, this can be achieved without the need for FSE to introduce any distributional assumptions on the error model. Therefore, FSE remains a completely distribution-free procedure, even when so extended.

In FSE, at any step the upper and lower bounds coalesce all the information that has been collected. The bounds are obtained by incrementally interacting choices together with model and functional assumptions, defining the area in which a representative shape would lie. Because of their informational role, a natural way to verify the goodness of FSE's output is to introduce additional questions to validate the bounds. Specifically, if the bounds are correct, a question that determines a point at or above (below) the upper (lower) bound should be answered with a preference for the sure amount (the lottery). This is a general approach that can be operationalized in several ways. To illustrate this idea, we here propose *FSE with validation* and *FSE with step-back*.

Arguably, the most immediate operationalization consists in validating the bounds obtained at termination. In this case, FSE can be extended to select and ask a user-defined number of questions that correspond to points on the final bounds. We suggest choosing these questions so as to maximize the Euclidian distance in the functional space from the points associated with previously asked questions. We refer to this extension of FSE as *FSE with validation*.

Alternatively, more than just the bounds at termination could be validated. Typically, this would make sense if we speculate that response errors might occur even for earlier questions. In this case, FSE can be extended to select questions corresponding to points not only on the bounds obtained at termination, but also on the bounds at a user-specified number of steps before termination. Analogously to FSE with validation, the number of questions per step-back can be user-specified, and these new questions can be chosen to be the furthest from the previous questions in the same sense as above. We refer to this extension of FSE as *FSE with step-back*.

Once we have collected additional choices to validate the bounds, these choices may be consistent or inconsistent with the preferences implied by the bounds. If the choices are consistent, the bounds are confirmed and a representative shape can be selected as usual, as the optimal separating hyperplane. If they are inconsistent, some of the initial and additional choices could be contradictory under the model assumptions. Pictorially, we could think of this as the admissible area illustrated, for instance, in Figures 4b not existing, at least over some parts of the support. In this case, the feasible space for the optimization programs underlying the bounds would be empty, and consequently, neither bounds nor a separating hyperplane can be determined. However, the core FSE ideas of bounds and a separating hyperplane can be re-introduced by considering the optimal hyperplane for non-separable data, i.e., the soft-margin support vector machine (SVM, Vapnik 1996). For contradictory choices, a representative shape that minimizes the number of misclassified choices can be selected using a soft-margin SVM. Based on this shape, FSE can attempt to classify choices as correct or as response errors and thus handle response errors. The SVM establishes a margin around this shape, determining an area where there is uncertainty regarding how to classify choices.[1] This idea is closely related to the bounds of FSE. The space between the bounds is the space for which the procedure is uncertain, in the sense that it has no information. The space between the margins is similarly uncertain, since the collected choices are contradictory and thus provide no definitive information.

FSE with validation and FSE with step-back are illustrated in Online Appendix OA.C.2subsection.3.2 and tested in a simulation study in Online Appendix OA.C.3subsection.3.3. Before we present these, we want to highlight that one could think of many alternative error handling schemes for FSE. For instance, one could repeat the elicitation twice and then find the shape that accommodates the largest number of preferences. For $\epsilon = .1$, this would still require fewer questions than bisection on average (21 against 25, according to our empirical results). Choosing a set of random validation questions would also be possible. The key aspect to note here is that, while FSE was developed with speed as a goal, its

[1]This idea further confirms the potential to link FSE with the concepts of choice inconsistency, as hinted to in §3.1.

core ideas can naturally be extended to accommodate inconsistent answers. The optimal separating hyperplane (hard-margin SVM) simply needs to be replaced by a soft-margin SVM. This will enable the procedure to attempt to spot response errors. Albeit with nuances, the bounds obtained with SVM still convey the idea of an area of unknown or uncertain preferences. Ultimately, devising and testing the best error-handling extension of FSE deserves a standalone project. The ideas presented here exemplify that this is possible using the machinery of FSE or ideas tightly connected to it.

OA.C.2 Illustration of error handling extensions

To illustrate FSE with validation and with step back, we introduce response error in the elicitation of the probability weighting function of Abdellaoui (2000, Table 9) shown in §4.1. In particular, for the purpose of illustration in this section, we fix all preferences expressed by the simulated respondent to be truthful till the eighth included; from the ninth until FSE terminates, all response are set to be errors; then, in the error handling part, all preferences are again truthful. Once FSE terminates, both validation or step back error handling use three additional questions. In a real experimental setting, these questions would integrate seamlessly with the elicitation, and the respondent would not perceive a discontinuity.

Figure OA.17figure.17 and OA.18figure.18 present how the step-by-step workings of FSE illustrated in Figure 6 change with validation and step back, respectively. Figure OA.19figure.19 and OA.20figure.20 present the change in the final output, analogously to Figure 7. In these graphs, erroneous responses are marked with an ex, to distinguish them from correct responses, indicated as usual by cycles. In Figures OA.17figure.17 and OA.18figure.18, the first eight steps are no different from the ones in Figure 6. The ensuing steps until termination are different because of response error. This leads FSE reaching termination at step 12. The termination step 12 has the ambivalent role of the final step of the elicitation and of the starting step of the error handling addendum. To make this perspicuous, we have doubled the illustration of step 12, with the second instance highlighting the bounds and questions used in the addendum. From the second illustration of the termination step onwards we are omitting the elicitation questions, for both clarity and to foreground that all the information used in error handling is contained in the bounds. From termination onwards the bounds are no longer updated and, consistently, the new preferences are no longer interacted with the feasible space: the selection of the representative shape is performed using SVM.

Let us first focus on the specificities of the two error handling alternatives, starting with validation as in Figure OA.17figure.17 and OA.19figure.19. As it can be seen in Figure OA.17figure.17, step 12, at step zero of validation the procedure chooses three additional questions, all belonging to the final bounds. The third question is a preference for a lottery and falls on the elicited upper bound. As such, it challenges the previously collected information, and leads a selected parametrization that is different from the one we would have gotten without the error handling addendum. This can be observed in Figure OA.19figure.19. Specifically, Figure OA.19asubfigure.19.1 compares the functional estimate that basic FSE would have produced (black dashed line) with the one we obtain when adding validation (orange full line). SVM tries to minimize the penalty for misclassification of preferences, leading to an estimated shape that takes smaller [larger] values in the left [middle] part of its support. This improves the quality of the recovered shape, as can be seen in Figure OA.19csubfigure.19.3. The parameters associated with these shapes are portrayed in Figure OA.19bsubfigure.19.2. Note that the validation estimate still belongs to the elicited feasible space. However, this will not always be the case.

Looking at FSE with step back now, we can note the following differences. With step back, the procedure chooses three questions each belonging to the bounds at and before termination. This can be seen in Figure OA.18figure.18, again from step 12. In this specific case of step back, none of the preferences collected in steps 13-15 dispute previous preferences. As a consequence, the estimated representative shape does not change compared to what basic FSE would have returned. This identity can be clearly observed in Figure OA.20figure.20. On one hand, this consistency is a sanity check: if the error handling questions provide no new information, the estimate should not change. On the other, it pinpoints the fact that the extensions presented here have the goal of showcasing how the machinery of FSE is readily amenable to handle errors. Devising a performing extension is worthy of a future project.

OA.C.3 Simulation study

We test the effectiveness of the proposed error handling extensions in a simulation study in which computerized respondent are susceptible to make erroneous choices, similarly to what we did in Appendix B.3. The simulated respondents obey RDU with linear utility. Consequently, their valuations depend on their individual pwfs. All pwfs are assumed to have the I-spline functional form described in §3.3. For each simulated respondent, an individual parametrization of this function is randomly chosen. Two random schemes are used: i) drawing from a uniform Dirichlet distribution and ii) bootstrapping the estimates of the experiment in §4.2. The first scheme uniformly explores all shapes that are possible given the I-spline functional form, which does not favor shapes displaying empirical regularities (typically, inverse S-shaped weighting). To visualize this, it can be thought of as uniformly sampling the admissible functional space at step 0 in Figure 6. The second scheme generates profiles that have a stronger connection with the empirical regularities we recovered in our experimental validation (§4.2), which align with the broader literature.

Our simulated respondents can incur response errors according to the two error models introduced in Appendix B.3: the random utility model and the constant error model. For the standard deviation of random utility, we consider multiple possible values: 0, .01, .025, .05, .1, .2, and .3. For the probability of constant error, we consider the values 0, .01, .025, .05, .1, .2, and .3, i.e., the same values used for random utility. Combining two random schemes with two error models, each with seven values for the error parameter, yields 28 possible simulation environments. We simulated $10,000$ respondents per simulation environment. Each respondent undertook FSE, FSE with validation, and FSE with step-back, with both $\epsilon = .1$ and $\epsilon = .05$. For both validation and step-back, we consider two variants: one asking three additional questions and one asking five. This means that, with three additional questions, FSE with validation will select three more questions belonging to the final bounds, while FSE with step-back will consider the bounds used in the last three questions of the elicitation and select an additional question that determines a point belonging to each of these bounds.

We evaluate robustness to error based on the accuracy of the elicitation of the three key elements associated with probability weighting, as captured by the functional parameters: the true possibility effect (λ_1), the true probabilistic sensitivity

($\lambda_2 + \lambda_3 + \lambda_4$), and the true certainty effect (λ_5). These elements capture the overall modeled behavior (the shape), as well as the most distinctive and economically relevant individual features. For brevity, here we present only the results for FSE with $\epsilon = .1$ and its extensions with 3-question and 5-question validation. The insights in the other conditions are qualitatively the same. Accuracy is assessed both as a point measure and in distribution.

For a point measure, we adopt the mean absolute distance of the estimates from the underlying truth (Mean Absolute Error; MAE).[2] To clarify the magnitude of the measurement error, we present these metrics relative to the average value of the true underlying parameter (relative MAE). These are illustrated in Figure OA.21figure.21 and can be interpreted as the expected percentage error of our estimate (in words, we can expect our estimate of (for instance) the possibility effect to be off by x%). Specifically, each column in Figure OA.21figure.21 corresponds to a combination of a random scheme and an error model, with the x-axis in each plot reporting the values of the associated error parameter (standard deviation or error probability). Note that, for perspicuity, the spacing of the values on the x-axis does not respect proportionality.

To provide a more complete picture, we also report the kernel-smoothed distributions of the parameter recovery errors in Figures OA.22figure.22 and OA.23figure.23. Here we can see the spread and direction of the errors, which can point to any systematic pattern (systematic over- or under-estimation, for instance). In these figures, each column corresponds to one of the key elements associated with probability weighting. Each row corresponds to a different error value for the corresponding error model being portrayed in the figure. Color-coded dashed vertical lines represent the median error of each procedure. Grey dashed vertical lines are placed at deciles on the x-axis, to help convey the spread of the errors.

Overall, the results indicate that the validation scheme can help improve the accuracy of the elicitation, although only in a subset of simulation environments. When the response errors have constant probability, validation helps in high error regimes, particularly for the accuracy of possibility and certainty effects (Figure OA.21figure.21). In this case, validation helps avoid some large misestimations, as can be noted in the bottom rows of Figure OA.22figure.22. Large deviations would occur because many errors have been collected during the elicitation, and

[2]We also test mean squared and median absolute errors, with the results being analogous and thus omitted here for brevity.

in these cases asking additional questions is beneficial, since it is more likely to gather accurate answers than errors. In general, however, additional questions expose the procedure to a larger number of errors, as can be seen in the last row of Figure OA.21figure.21. Finally, note that going from 3 to 5 validation questions increases accuracy very limitedly.

While this simulation study demonstrates the potential to extend FSE to the handling of response errors, we feel that the key insight of these results is more general and aligns with the discussion in §5. When it comes to identifying response errors, validation questions are no free lunch: If our initial preferences are subject to response error, our validation questions will be as well. This makes it complicated, based on a small or moderate number of questions, to ask a learning model to accurately and confidently distinguish faithful preferences from response errors. We should also consider that in an experimental setting the likelihood of response errors might increase with the number of questions, again as discussed in §5. Ultimately these results confirm that, as experimenters, we may have a better chance of getting a more accurate representation of personal preferences with fewer questions that a respondent can answer with a high level of attention. FSE goes in this direction.

Figure OA.17: Step-by-step elicitation with validation - GE with $r = .84, s = .65$.

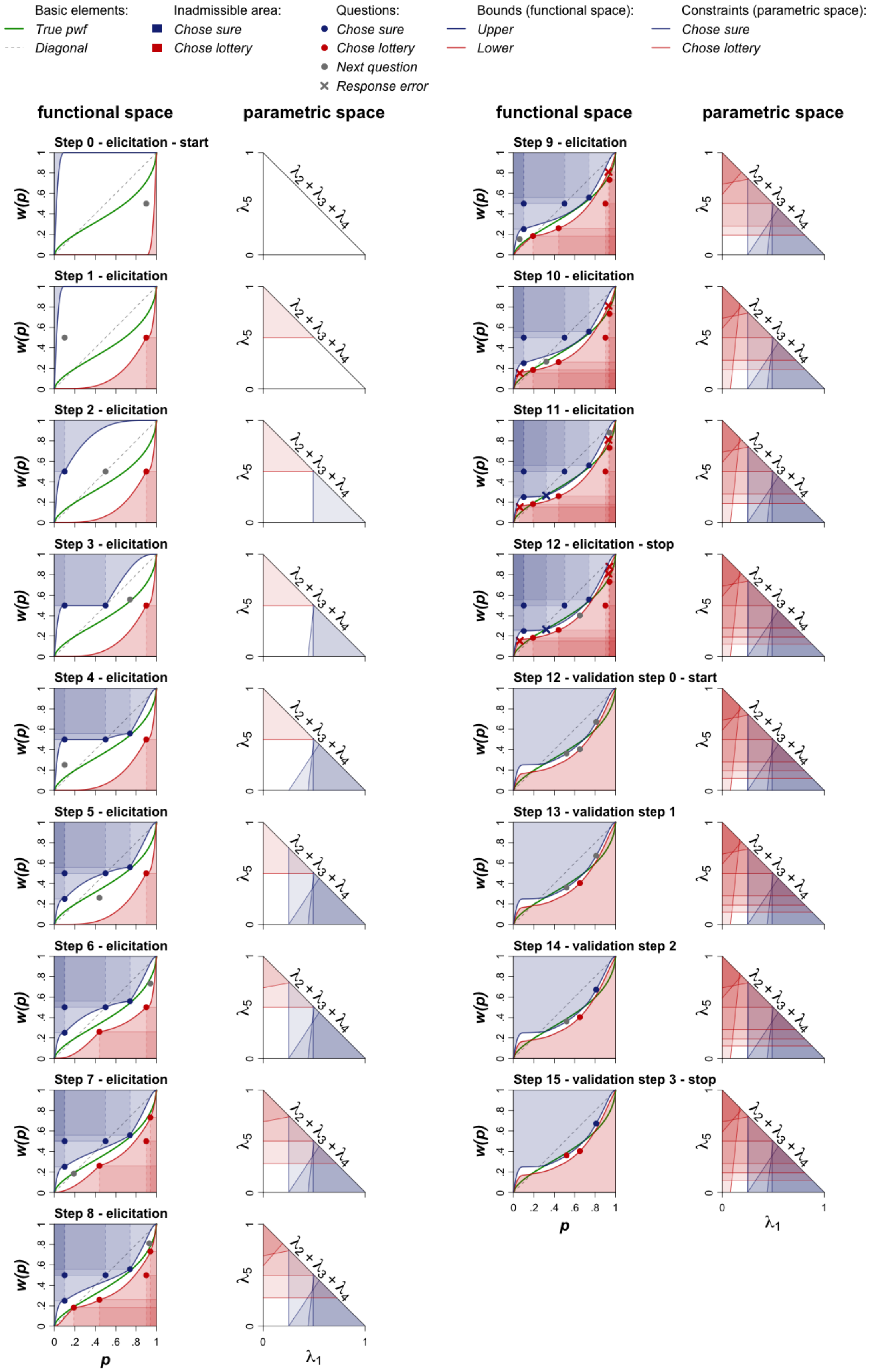


Figure OA.18: Step-by-step elicitation with step back - GE with $r = .84, s = .65$.

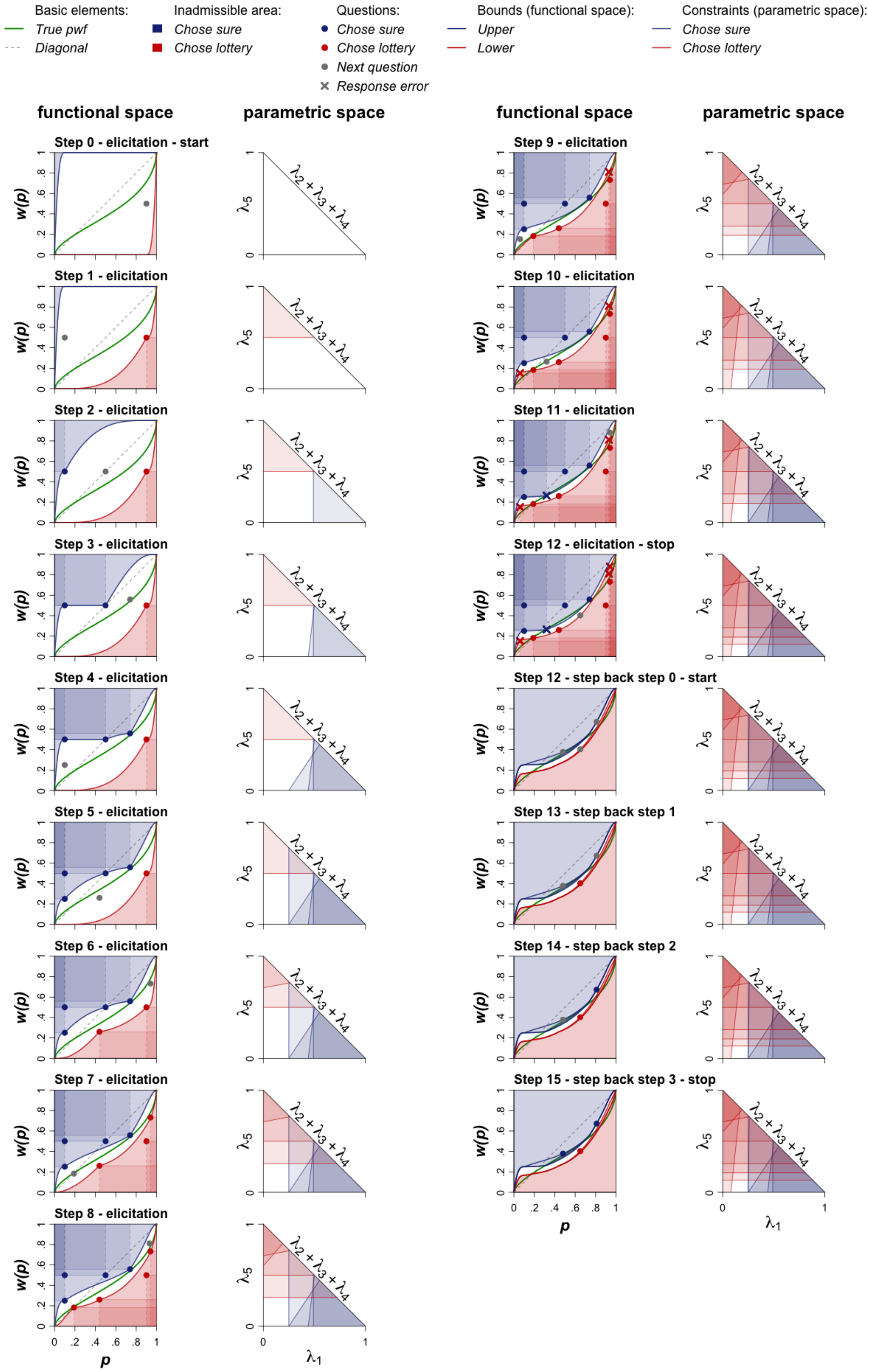


Figure OA.19: Final output with validation.

(a) functional space. (b) parametric space. (c) distances at termination.

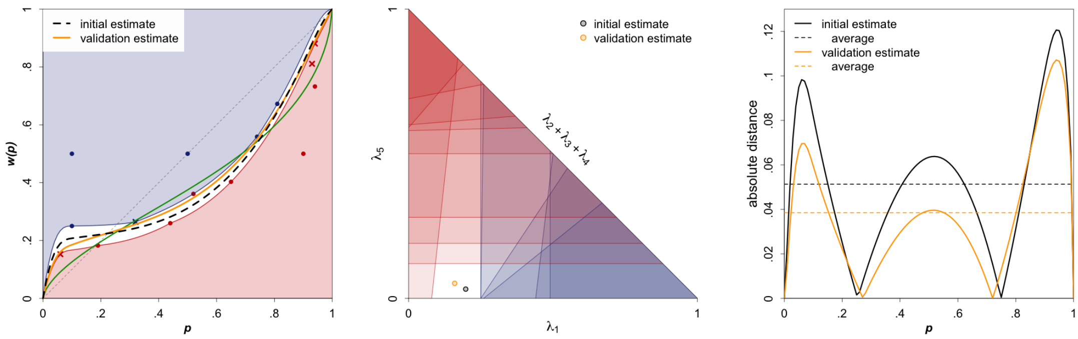


Figure OA.20: Final output with step back.

(a) functional space. (b) parametric space. (c) distances at termination.

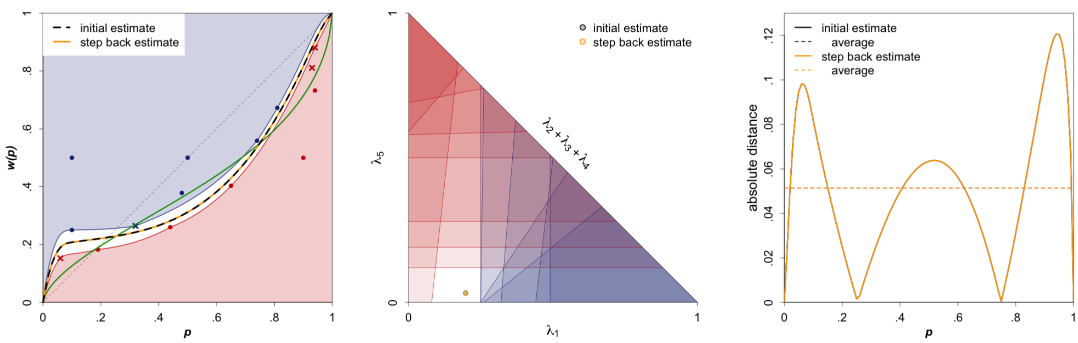


Figure OA.21: Robustness to error with $\epsilon = .1$ and validation.

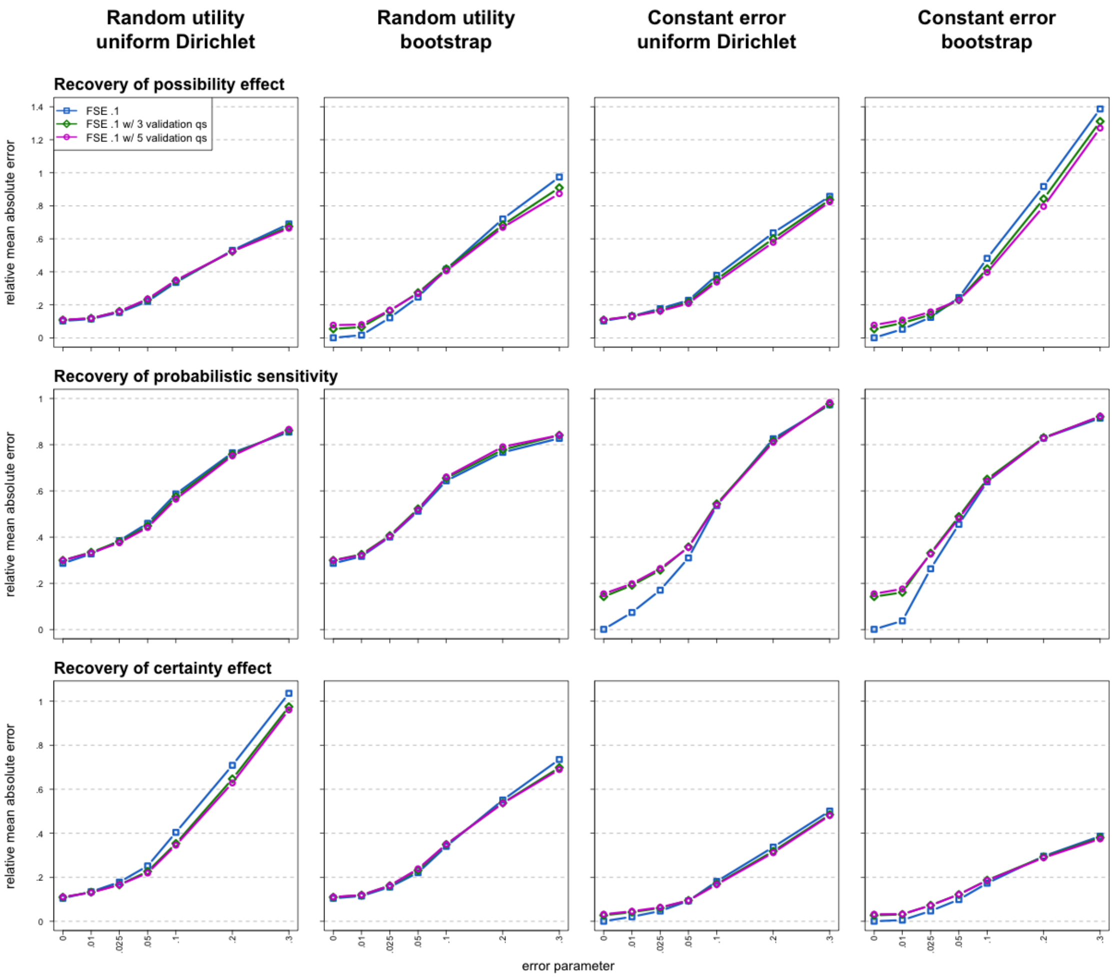


Figure OA.22: Distribution of estimation errors with uniform Dirichlet, random utility, $\epsilon = .1$, and validation.

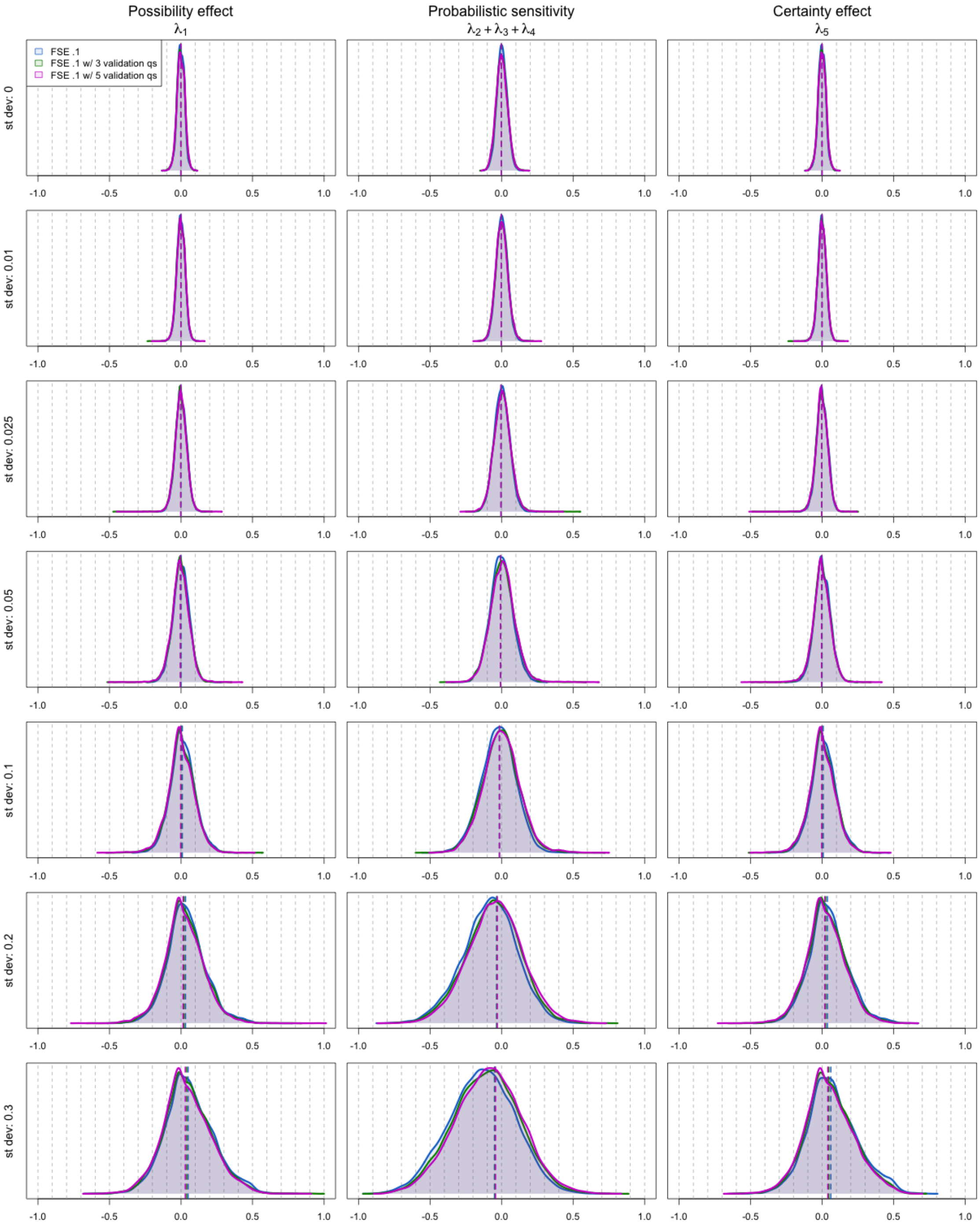
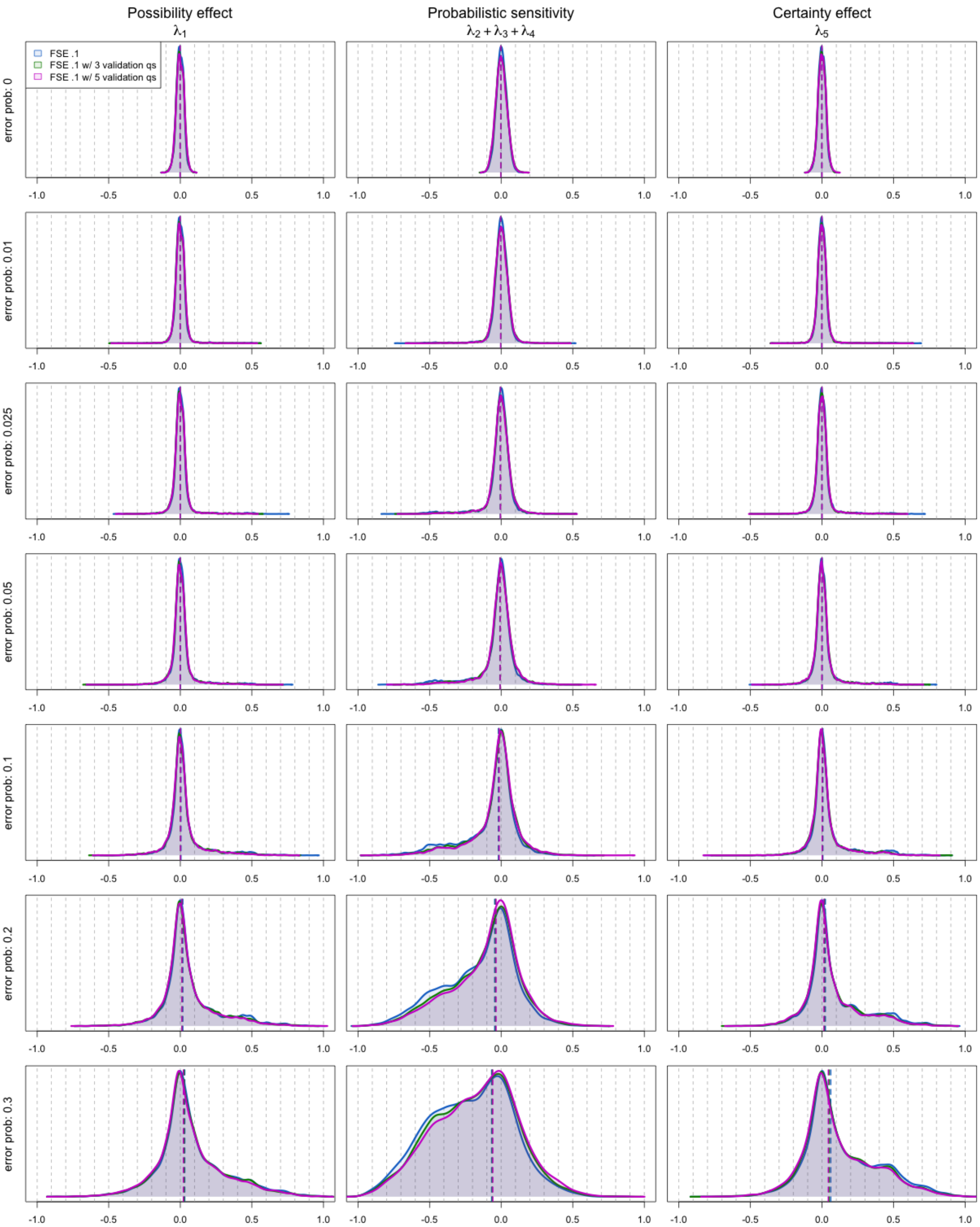


Figure OA.23: Distribution of estimation errors with uniform Dirichlet, constant error, $\epsilon = .1$, and validation.



OA.D Elicitation of the utility function

OA.D.1 The structure of FSE for EU

Expected Utility (EU henceforth) is the cornerstone model of decision theory (von Neumann and Morgenstern, 1944, 1947). In spite of numerous descriptive limitations, it is still the most popular choice for modeling individual behavior and studying risk attitude in economics, operations research, management science, and decision analysis. We here show how our procedure can be inflected to elicit utility with a low number of easy binary choices. After defining the procedure for EU, we will showcase its ability to recover a close approximation to a true underlying model by eliciting the utility function estimated by Holt and Laury (2002).

Under EU, a generic lottery $(x_1 : p_1 \; ; \; \dots \; ; \; x_n : p_n)$ writes as the probability weighted sum of utilities of the outcomes:

$$(x_1 : p_1 \, ; \, \dots \, ; \, x_n : p_n) \quad \mapsto \quad \sum_{i=1}^{n} p_i u(x_i), \tag{1}$$

where u is the utility function, dependence on the initial wealth level left implicit, and no ranking of the outcomes is assumed, as opposed to (1). The utility function is assumed to be continuous and non-decreasing in x. For binary lotteries, which are sufficient for our procedure, (1equation.1) simplifies to the following:

$$(x : p \, ; \, y) \quad \mapsto \quad pu(x) + (1-p)u(y), \tag{2}$$

without any assumptions on the ranking of the outcomes, conversely to (2). Since monotonicity holds for u under EU as it holds for w under RDU, the interpretation of a preference between $(x : p_i; y)$ and z_j can be interpreted similarly to (8) and (9). Under EU the interpretation is the following:

$$(x : p_i \, ; \, y) \succcurlyeq z_j \iff u(z) \leq p, \qquad \forall \, z \leq z_j, \, p \geq p_i, \tag{3}$$

$$z_j \succcurlyeq (x : p_i \, ; \, y) \iff u(z) \geq p, \qquad \forall \, z \geq z_j, \, p \leq p_i, \tag{4}$$

where we have normalized $u(y) = 0$ and $u(x) = 1$. The comparison between (8) and (9), and (3equation.3) and (4equation.4) foregrounds the duality between RDU with linear utility and EU.

Similarly to what we did in (A.4), we again choose to define the functional form for the utility function u, after the aforementioned normalization, to be a convex combination of basis I-splines. This choice respects the assumption on u. The utility function u therefore writes as:

$$u(z) = \sum_k \lambda_k I_k(z). \tag{5}$$

with $\lambda_k \geq 0$, $\forall k$ and $\sum_k \lambda_k = 1$. Note that, while for probability the support is naturally bounded to the interval $[0, 1]$, here the basis expansion is defined over outcome values, spanning $\mathbb{R}$. This is not a problem: an I-spline expansion can be defined over any closed interval $[L, U]$, as mentioned in §2.2. The most natural candidate for this interval when eliciting EU is the range of outcomes used in the experiment. In our experimental validation we used outcomes in $[10, 120]$. Had we elicited EU, this would have been the most natural support for defining the basis expansions.

For the functional choice of (5equation.5), (3equation.3) and (4equation.4) take the following linear form:

$$(x : p_i \,;\, y) \succcurlyeq z_j \iff \sum_k \lambda_k I_k(z) \leq p, \qquad \forall\, z \leq z_j,\, p \geq p_i, \tag{6}$$

$$z_j \succcurlyeq (x : p_i \,;\, y) \iff \sum_k \lambda_k I_k(z) \geq p, \qquad \forall\, z \geq z_j,\, p \leq p_i. \tag{7}$$

Again this is dual to (8) and (9), and entail the same advantages discussed in §3.3.

Based on the above, we can now write the sequencing rule to elicit the utility function under EU. As we did for the elicitation of the probability weighting function, at every step t our procedure will ask only binary choices between a sure amount z_t and a two-outcome lottery $(x : p_t \,;\, y)$. Outcomes x and y are set to be fixed throughout the procedure, to reduce complexity. As in the above, they are normalized to be $u(y) = 0$ and $u(x) = 1$. The basis expansion to define u as in (5equation.5) will then span the interval $[y, x]$, assuming $x \geq y$. This will ensure, under a convex combination of $\boldsymbol{\lambda}$ that u is monotonically non-decreasing, $u(y) = 0$, and $u(x) = 1$.

Again, at step t we seek the ordered pair (p_t, z_t) that would be optimally informative, according to our definition and given the previous $t - 1$ preferences. In our setup, this would be the ordered pair that bisects the maximal uncertainty regarding the function values, expressed as the maximal vertical distance between the bounds on the function. $\overline{\mathcal{U}}_t = \{\overline{u}_t(z), \forall z \in [y, x]\}$ and $\underline{\mathcal{U}}_t = \{\underline{u}_t(z), \forall z \in [y, x]\}$ are respectively the upper and lower bound on u at step t. As $\overline{w}_t(p)$ is defined in (A.1) under RDU with linear utility, $\overline{u}_t(z)$ is the solution to the following constrained optimization problem:

$$
\begin{aligned}
\overline{u}_t(z) \;=\; & \max_{u} && u(z), && \text{(8equation.8)} \\
& \text{subject to} && u(y) = 0, && \text{(8a)} \\
& && u(x) = 1, && \text{(8b)} \\
& && u(z) \leq u(q), \quad \forall\, z, q \in [y, x] : z \leq q, && \text{(8c)} \\
& && u(z_j) \geq p_j \qquad \forall j \in \{i : z_i \succcurlyeq (x : p_i\,;\; y), i \leq t, i \in \mathbb{N}\}, && \text{(8d)} \\
& && u(z_j) \leq p_j \qquad \forall j \in \{i : (x : p_i\,;\; y) \succcurlyeq z_i, i \leq t, i \in \mathbb{N}\}, && \text{(8e)}
\end{aligned}
$$

and $\underline{u}_t$ is the solution to the minimization problem under the same constraints.

Formally, we seek the dual of (A.2) and (A.3), which write as:

$$
z_t \;=\; \operatorname*{arg\,max}_{z \,\in\, [y,x]} \quad \overline{u}_{t-1}(z) - \underline{u}_{t-1}(z), \tag{9}
$$

$$
p_t \;=\; \frac{\overline{u}_{t-1}(z_t) + \underline{u}_{t-1}(z_t)}{2}. \tag{10}
$$

The considerations that apply to (A.1) and (A.2) equally apply to (8equation.8) and (9equation.9). As (A.1) can be turned into a linear program by assuming (A.4), similarly for the functional choice in (5equation.5) the optimization problem

(8equation.8) turns into the following linear program:

$$\overline{u}_t(z) = \max_{\boldsymbol{\lambda}} \quad \sum_k \lambda_k I_k(z), \tag{11equation.11}$$

$$\text{subject to} \quad \lambda_k \geq 0, \quad \forall\, k, \tag{11a}$$

$$\sum_k \lambda_k = 1, \tag{11b}$$

$$\sum_k \lambda_k I_k(p_j) \geq p_j \quad \forall j \in \{i : z_i \succcurlyeq (x : p_i\, ;\ y), i \leq t, i \in \mathbb{N}\}, \tag{11c}$$

$$\sum_k \lambda_k I_k(p_j) \leq p_j \quad \forall j \in \{i : (x : p_i\, ;\ y) \succcurlyeq z_i, i \leq t, i \in \mathbb{N}\}. \tag{11d}$$

Further, again imposing a discretization such that $z \in \mathcal{Z} = \{z_1, \ldots, z_n : z_i \in [y, x]\}$ makes (9equation.9) solvable. A representative function for each subject can be chosen using the same technique presented in §A.4 for the case of the probability weighting function. Our procedure for eliciting the utility function under EU is summarized in Algorithm 1algocf.1.

OA.D.2 Simulation study for EU

We showcase the ability of FSE to elicit the utility function under EU for a simulated respondent. The respondent's utility function is set to the specification and parameter estimates of Holt and Laury (2002): It is an expo-power form with $\alpha = .029$ and $r = .269$. The respondent provides accurate answers in accordance with EU maximization. We align the interval of outcomes used by FSE with the incentives of Holt and Laury (2002, Table 1): The interval $[y, x]$ is set to $[.1, 3.85]$ (currency is irrelevant in the simulation). We set $\boldsymbol{\xi} = (.1, 2.975, 3.85)$; this coarser partition is consistent with the generally smoother shape of the utility function compared to the pwf. We discretize z to increments of .05; ϵ is again fixed to .1.

The workings of FSE are illustrated in Figure OA.24figure.24. In comparison to Figure 6, note the changes in the axis, reflecting the elicitation of utility. More specifically, we bound normalized utility (y-axis from 0 to 1) over the outcome interval (x-axis from .10 to 3.85). Compared to Figure 6, a larger portion of the space is excluded ex ante at step 0, reflecting the use of a coarser $\boldsymbol{\xi}$. Also note that the first question is chosen to be the same question used for the elicitation

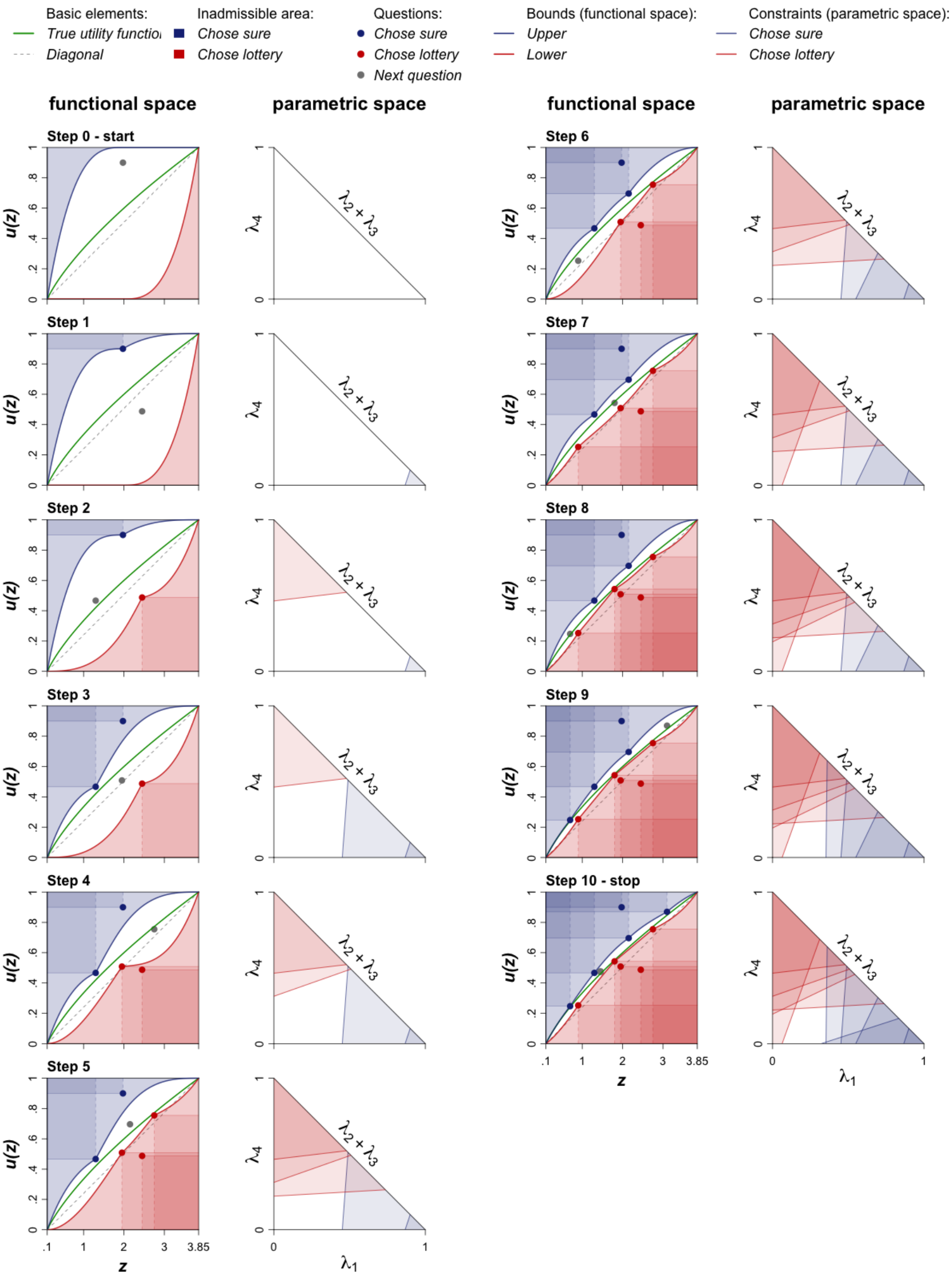
Algorithm 1: Fast and simple elicitation of u under Expected Utility.

Input: x, y, b, $\boldsymbol{\xi}$, ϵ, $\mathcal{Z}$
Output: $\hat{\boldsymbol{\lambda}}$
begin
    $t \leftarrow 0$;
    `Update` $\overline{\mathcal{U}}_t$, $\underline{\mathcal{U}}_t$;
    while $\max\limits_{z \in \mathcal{Z}}\{\overline{u}_t(z) - \underline{u}_t(z)\} > \epsilon$ **do**
        $t \leftarrow t + 1$;
        $z_t \leftarrow \arg\max\limits_{z \in [y,x]} \overline{u}_{t-1}(z) - \underline{u}_{t-1}(z)$;
        $p_t \leftarrow \frac{\overline{u}_{t-1}(z_t)+\underline{u}_{t-1}(z_t)}{2}$;
        `Ask`: $z_t \overset{?}{\succcurlyeq} (x : p_t\,;\ y)$;
        if $z_t \succcurlyeq (x : p_t\,;\ y)$ **then**
            $s_t \leftarrow 1$;
        else
            $s_t \leftarrow 0$;
        end
        `Update` $\overline{\mathcal{U}}_t$, $\underline{\mathcal{U}}_t$;
    end
    `SelectFunction`
end

of the pwf. The location in this figure is shifted across the diagonal, compared to Figure 6, reflecting the duality with the elicitation in §4. Analogously to what could be observed in Figure 6, as the simulated respondent makes binary choices, FSE incrementally chisels out parts of the admissible outcome-utility space. After 10 questions, the termination condition is met, with the maximal distance between the bounds having been reduced below .1 over the entire support.

At termination, FSE can select a representative utility function. The results are illustrated in Figure OA.25figure.25. This figure is analogous to Figure 7. In the functional space, it can again be seen that the selected function is a close approximation to the truth. The final distance between the bounds is again, and by construction, below the threshold $\epsilon = .1$ over the entire discretized support. This distance is even smaller than in the case of §4.1: It is about .07. The recovery of the true utility function is accurate: The distance of the estimate is on average .018. Thus, FSE has faithfully elicited the true utility function.

Figure OA.24: Step-by-step illustration of FSE for the utility function.

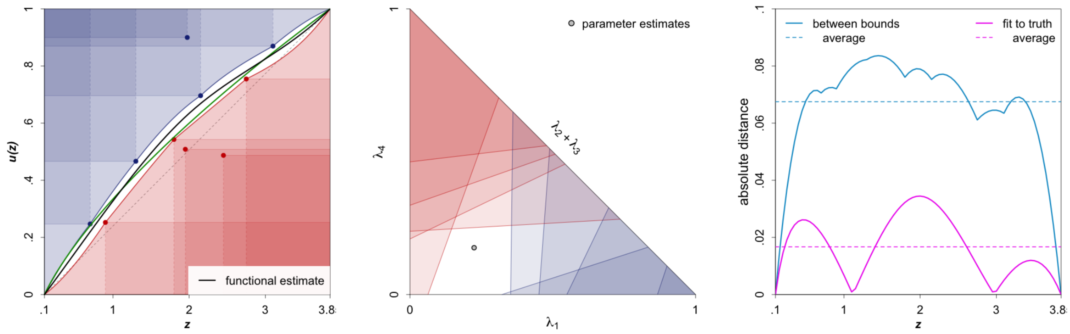


OA.E FSE with tradeoff method

This section presents a supplemental study that was conducted as part of the online experiment discussed in §4.3 and Appendix B.4. This study conjugates the estimation of the probability weighting function using FSE with the elicitation of the utility function using the tradeoff method (Wakker and Deneffe, 1996). It

Figure OA.25: Final output of FSE for the utility function.

(a) functional space. (b) parametric space. (c) distances at termination.



probes the plausibility of the assumption of linear utility given the chosen stimuli, and it showcases how FSE can be used *as is* in the larger context of the elicitation of two decision-theoretic functions.

OA.E.1 Details of the experiment

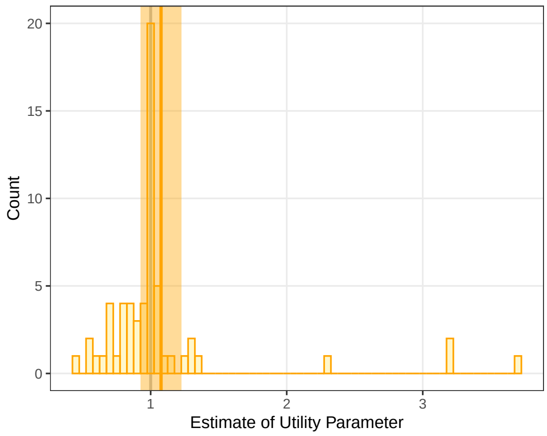
As per our preregistration, https://aspredicted.org/gjgz-ngnq.pdf, our online experiment included a supplemental part in which 60 participants also underwent the elicitation of utility. For this, the first task in the experiment was the tradeoff procedure of Wakker and Deneffe (1996). The obtained utility function was then input in the elicitation of the probability weighting function using FSE, before presenting the participants with the out-of-sample questions. We here only discuss the details that differ from §4.3 and Appendix B.4.1.

For this supplemental investigation, we recruited an additional 60 online participants on Prolific. All participants were from the US, their primary language is English, and they had not been part of the main experiment. They all completed the study without encountering technical problems, yielding a final sample of 60. Anticipating that the experiment would take slightly longer due to the additional tradeoff task, we adjusted the participation fee to \$2.2 (£1.80). Bonus payments via a random incentive scheme were implemented consistently with the main experiment. Two participants were drawn and received an additional \$24 in total.

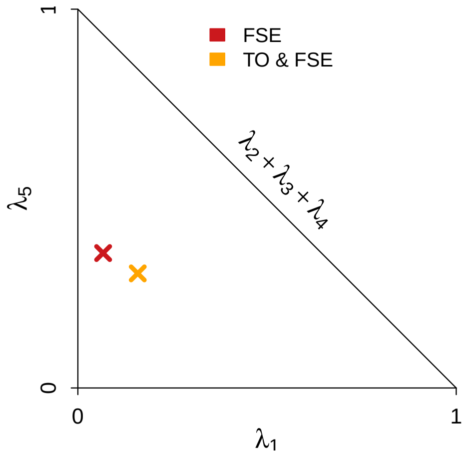
The tradeoff task asks the participant to choose between two lotteries $(x_j : p;\ R)$ and $(x_{j+1} : p;\ r)$, where r and R are the two fixed gauge amounts with $r < R$, x_0 and x_1 are susceptible to change after each choice, and $j \in \{1, \ldots, J\}$.

Figure OA.26: Estimated parameters with and without the tradeoff task.

(a) Utility.



(b) Probability weighting.



As per common practice, we set $p = \frac{1}{3}$, since probability distortion is considered to be small at this level. We choose $x_1 = \$2$ and $R = \$18$, for consistency with the lottery used in the elicitation of the pwf. Given these choices, we take $r = 15$ and the initial value of x_{j+1} is chosen to equate the two lotteries in expected value, i.e. in the first choice $x_2 = \$8$. After each choice, the value of x_{j+1} is either increased by \$1 if $(x_j : p;\ R)$ is chosen, or decreased by \$1 otherwise. When the participant changes the choice from one lottery to the other, one last refinement step is presented, in which the value of x_{j+1} is adjusted by \$.5. For a given value of x_j, the task presents at most 5 choices. After the refinement step or after the fifth choice, the final value of x_{j+1} is saved and j is increased by one. In our study, the tradeoff task collects four values, i.e. $J = 4$. The utility function is assumed to have a power form such that $u(z) = z^\gamma$. The utility parameter γ that best fits the four collected values is estimated using least squares. Since least squares estimation can be conducted instantaneously at this scale, there is no discontinuity between the tradeoff task and FSE.

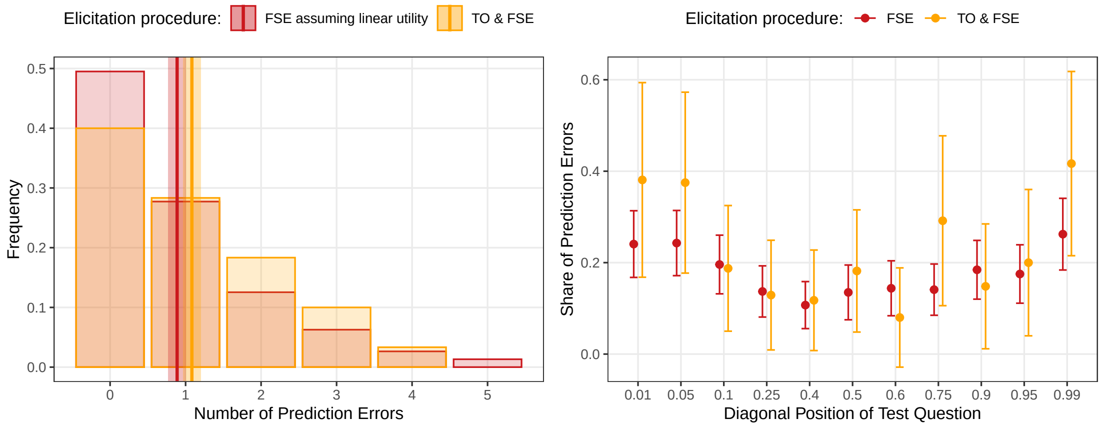
OA.E.2 Results of the experiment

Our primary interest is to probe the viability of our assumption of linear utility. For our chosen utility function, linearity is indicated by $\gamma = 1$.[3] The distribution of the estimates of γ is presented in Figure OA.26asubfigure.26.1 and decisively

[3] While the tradeoff method accommodates non-parametric analysis, for brevity and clarity we here focus on the parametric results. The assumed power function is widely used and favors comparability with previous studies (Bleichrodt et al., 2010).

Figure OA.27: Prediction accuracy with and without the tradeoff task.

(a) Overall accuracy. (b) Accuracy per question.



indicates that linear utility is a plausible assumption. Notably, for one in three subjects the estimated parameters is in $[.975, 1.025]$ and for more than one in two it is in $[.9, 1.1]$; the average estimate of 1.07 is not distinguishable from 1 at the conventional significance level (p-value = .32 in two-sided t-test).

When allowing non-linear utility, the parameter estimates for probability weighting are partially affected, as shown in Figure OA.26bsubfigure.26.2. In this case, we record larger possibility effect (p-value = .001 in one-sided t-test), whereas the other changes are not significant. It is compelling to note that the inclusion of the utility transformation does not seem to improve out-of-sample precision. This can be seen in Figures OA.27asubfigure.27.1 and OA.27bsubfigure.27.2, which are analogous to Figures 10 and 12a. The results generally point to a decrease in accuracy when the utility transformation is included, however none of these changes is significant at the conventional level. In light of the observations of §5, a decrease in predictive accuracy when including the tradeoff task could be expected: attention seems to diminish in the number of questions and the out-of-sample task is the last part of the experiment.

References

Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science*, 46(11):1497–1512.

Bateman, I., Dent, S., Peters, E., Slovic, P., and Starmer, C. (2007). The affect heuristic and the attractiveness of simple gambles. *Journal of Behavioral Decision Making*, 20(4):365–380.

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review*, 115(2):463.

Bleichrodt, H., Cillo, A., and Diecidue, E. (2010). A quantitative measurement of regret theory. *Management Science*, 56(1):161–175.

Cavagnaro, D. R., Gonzalez, R., Myung, J. I., and Pitt, M. A. (2013a). Optimal decision stimuli for risky choice experiments: An adaptive approach. *Management Science*, 59(2):358–375.

Cavagnaro, D. R., Pitt, M. A., Gonzalez, R., and Myung, J. I. (2013b). Discriminating among probability weighting functions using adaptive design optimization. *Journal of Risk and Uncertainty*, 47(3):255–289.

Cubitt, R. P., Starmer, C., and Sugden, R. (1998). On the validity of the random lottery incentive system. *Experimental Economics*, 1(2):115–131.

Diecidue, E., Levy, M., and van de Ven, J. (2015). No aspiration to win? an experimental test of the aspiration level model. *Journal of Risk and Uncertainty*, 51(3):245–266.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655.

Lopes, L. L. and Oden, G. C. (1999). The role of aspiration level in risky choice: A comparison of cumulative prospect theory and sp/a theory. *Journal of mathematical psychology*, 43(2):286–313.

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

Vapnik, V. (1996). *The Nature of Statistical Learning Theory*. Springer.

von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

von Neumann, J. and Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton University Press.

Wakker, P. and Deneffe, D. (1996). Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Management Science*, 42(8):1131–1150.