Roadmap

Adaptive Experimental Design

Changkuk Im

OSU Theory/Experimental Reading Group
April 23, 2025

Why Adaptive Design?

1. To better compare different treatments
   - Exley and Kessler (2024); Exley (2020)

2. To obtain precise model predictions
   - Model comparison: Halevy et al. (2018); Zrill (2024); Somerville (2022); Im (2024)
   - Test predictions: Toney et al. (2023)

Accurate Elicitation is HARD!

Exley and Kessler (2024)

Q. How DMs respond to a **zero** being added to a payoff?

- 150 vs $51 + 51 + 51 + 51$
- 150 vs $51 + 51 + 51 + 51 + 0$

Charity/Charity treatment

- 150 cents to the Make-A-Wish Foundation
- Sum of four or five summands to the Make-A-Wish Foundation

Self/Charity treatment

- X cents to oneself
- Sum of four or five summands to the Make-A-Wish Foundation

Accurate Elicitation is HARD!

Exley and Kessler (2024)

Use MPL to elicit X where

- 150 cents to the Make-A-Wish Foundation
- X cents to oneself

are indifferent

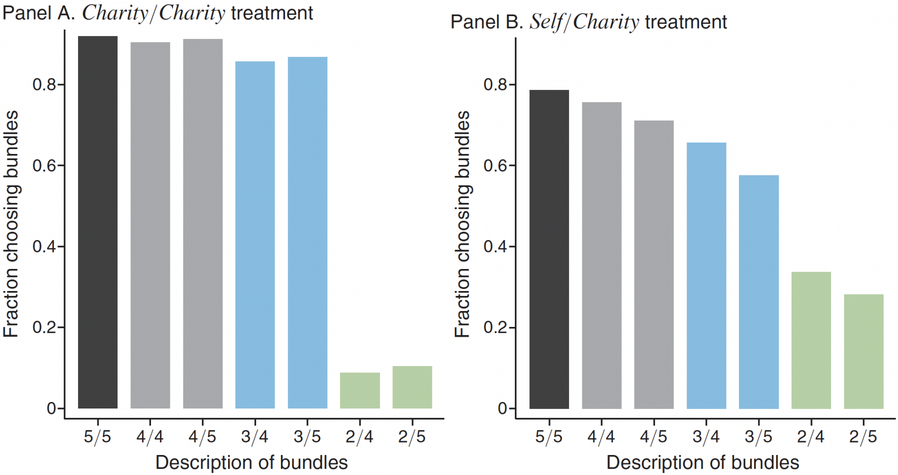
Please indicate which payment option you prefer on each row by clicking on the row where you would like to switch from choosing the option on the left to choosing the option on the right.

(Note that you cannot click on the submit button until you have selected an answer.)

DONATION FOR MAKE-A-WISH FOUNDATION		BONUS PAYMENT FOR YOU
150 CENTS	OR	0 CENTS
150 CENTS	OR	5 CENTS
150 CENTS	OR	10 CENTS
150 CENTS	OR	15 CENTS
150 CENTS	OR	20 CENTS
150 CENTS	OR	25 CENTS
150 CENTS	OR	30 CENTS
150 CENTS	OR	35 CENTS
150 CENTS	OR	40 CENTS
150 CENTS	OR	45 CENTS
150 CENTS	OR	50 CENTS
150 CENTS	OR	55 CENTS
150 CENTS	OR	60 CENTS
150 CENTS	OR	65 CENTS
150 CENTS	OR	70 CENTS
150 CENTS	OR	75 CENTS
150 CENTS	OR	80 CENTS
150 CENTS	OR	85 CENTS
150 CENTS	OR	90 CENTS
150 CENTS	OR	95 CENTS
150 CENTS	OR	100 CENTS
150 CENTS	OR	105 CENTS
150 CENTS	OR	110 CENTS
150 CENTS	OR	115 CENTS
150 CENTS	OR	120 CENTS
150 CENTS	OR	125 CENTS
150 CENTS	OR	130 CENTS
150 CENTS	OR	135 CENTS
150 CENTS	OR	140 CENTS
150 CENTS	OR	145 CENTS
150 CENTS	OR	150 CENTS

Accurate Elicitation is HARD!

Exley and Kessler (2024)



Precise Model Predictions

Halevy et al. (2018)

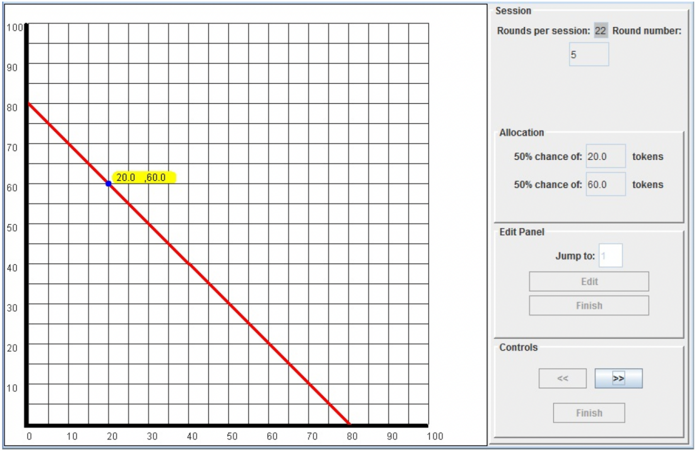
Q. Which estimation method has better out-of-sample predictions?

Experimental Design

1. Linear budget sets
   - Choose portfolio $x^i = (x_1^i, x_2^i)$ from a linear budget (22 budgets)
   - Estimate model parameters (disappointment aversion w/ CRRA) using two different methods (NLLS vs MMI)
2. Binary choices
   - Generate 9 binary choices (safe vs risk) where the predictions from NLLS and MMI differ ▸ Generating rule
   - The interface is similar to that in Task 1

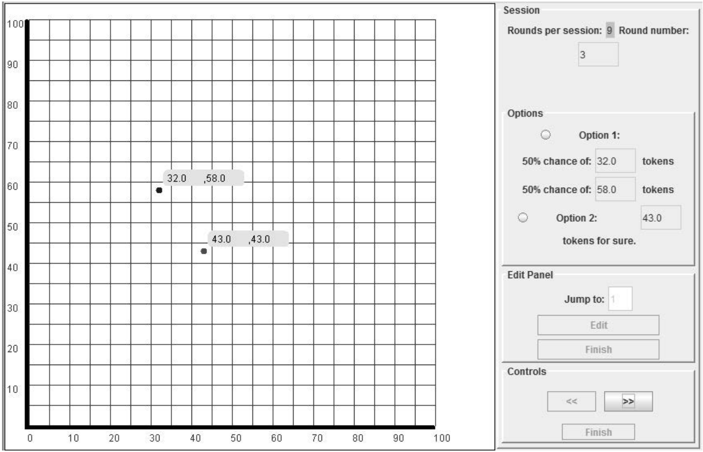
Precise Model Predictions

Halevy et al. (2018)



Precise Model Predictions

Halevy et al. (2018)



Precise Model Predictions

Halevy et al. (2018)

Main result

- 986/1827 (54%) choices are consistent with the predictions by MMI
  - Hard to verify whether the estimations from the linear budget task are valid for the choice task
  - Still, they tried to make the two tasks relevant by making the interfaces similar

Evaluation Task is Not Perfect

- Gap between the evaluation and choice tasks
  - Prediction accuracy of evaluation methods in binary choices (Hascher et al., 2021): WTP (71.8%), BDM (71.3%), unincentivized rating (73.6–78.4%)

E.g. Errors may play different roles (McGranaghan et al., 2024b)

Evaluation Task is Not Perfect

- Gap between the evaluation and choice tasks
  - Prediction accuracy of evaluation methods in binary choices (Hascher et al., 2021): WTP (71.8%), BDM (71.3%), unincentivized rating (73.6–78.4%)

  E.g. Errors may play different roles (McGranaghan et al., 2024b)

- Low correlation across elicitation methods
  - Risk (Charness et al., 2013; Crosetto and Filippin, 2016; Zhou and Hey, 2018)
  - Framing (Beauchamp et al., 2020; Brown and Healy, 2018; Sprenger, 2015)

Evaluation Task is Not Perfect

- Gap between the evaluation and choice tasks
  - Prediction accuracy of evaluation methods in binary choices (Hascher et al., 2021): WTP (71.8%), BDM (71.3%), unincentivized rating (73.6–78.4%)

  E.g. Errors may play different roles (McGranaghan et al., 2024b)
- Low correlation across elicitation methods
  - Risk (Charness et al., 2013; Crosetto and Filippin, 2016; Zhou and Hey, 2018)
  - Framing (Beauchamp et al., 2020; Brown and Healy, 2018; Sprenger, 2015)
- Elicitation may be susceptible to errors
  - McGranaghan et al. (2024a) measure the values of lotteries multiple times: The correlation coefficients are 0.254–0.696

▸ Table ▸ Screenshot

How to Do Better?

- Gap between the evaluation and choice tasks

- Low correlation across elicitation methods

- Elicitation may be susceptible to errors

How to Do Better?

- Gap between the evaluation and choice tasks
  - Make the interfaces as similar as possible (Halevy et al., 2018; Zrill, 2024)
  - Complement analysis by using comparative statics (Exley and Kessler, 2024)

- Low correlation across elicitation methods

- Elicitation may be susceptible to errors

How to Do Better?

- Gap between the evaluation and choice tasks
  - Make the interfaces as similar as possible (Halevy et al., 2018; Zrill, 2024)
  - Complement analysis by using comparative statics (Exley and Kessler, 2024)

- Low correlation across elicitation methods
  - Choose a method that matches with the context of the main task (Im, 2024; Zhou and Hey, 2018)

- Elicitation may be susceptible to errors

How to Do Better?

- Gap between the evaluation and choice tasks
  - Make the interfaces as similar as possible (Halevy et al., 2018; Zrill, 2024)
  - Complement analysis by using comparative statics (Exley and Kessler, 2024)

- Low correlation across elicitation methods
  - Choose a method that matches with the context of the main task (Im, 2024; Zhou and Hey, 2018)

- Elicitation may be susceptible to errors
  - Choose a method that is most reliable
  - Use averaged values (Gillen et al., 2019)
  - Use adaptive algorithms: DOSE (Chapman et al., 2024) ▸ More algorithms

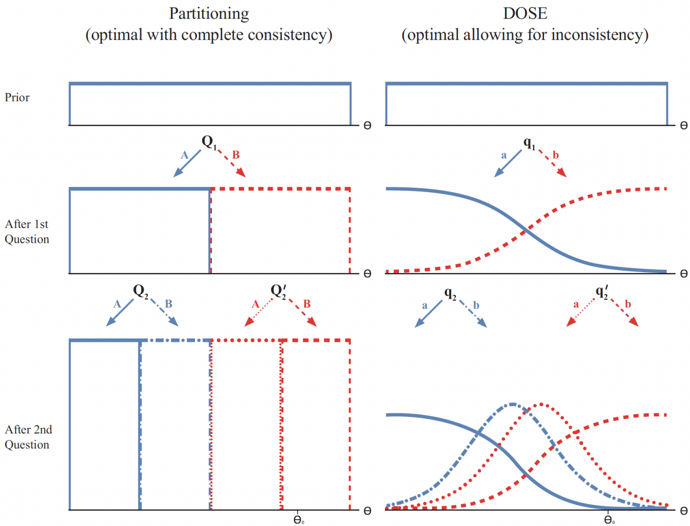
Dynamically Optimized Sequential Experimentation (DOSE)

Briefly description

- Structural model (parametric specification)
  - Prospect Theory w/ power utility (risk and loss aversion)
- Prior distribution over parameters
  - Uniform distribution
- Set of choices
  - Binary choice
- How parameters map to choices
  - Logit function (error can be taken into account)
- Information criterion
  - Expected Kullback-Leibler divergence between the prior and possible posteriors

$\Longrightarrow$ Accurate estimation w/ a few questions and taking into account errors

Partitioning vs DOSE



Another Example

Toney et al. (2023)

Q. Does calorie information affect choices?

Experimental design

1. Elicit WTP
   - 600-calorie Sandwich, 300/600/900-calorie wraps
   - Calorie information is not revealed
2. Binary choices close to indifferent by monetary compensation
   - Control: Make choices w/o calorie information
   - Calorie information: Make choices w/ calorie information

Another Example

Toney et al. (2023)

Chicken Sandwich

Ingredients: Grilled, 100% Antibiotic-Free Chicken Breast, Leafy Lettuce, Tomato, Swiss Cheese, Homemade Ranch Dressing between Two Multigrain Wheat Bread.

Note: You will be required to eat at least some of the food selected for implementation during today's session, so consider the options carefully.

With a budget of $10, what is the maximum amount that you are willing to pay for the meal shown above?

Amount ($)

0 1 2 3 4 5 6 7 8 9 10

Response

Another Example

Toney et al. (2023)

Option	A	B
Image		
Food Item	Chicken Sandwich	Chicken Caesar Wrap *B*
Description	Grilled, 100% Antibiotic-Free Chicken Breast, Leafy Lettuce, Tomato, Swiss Cheese, Homemade Ranch Dressing between Two Multigrain Wheat Bread.	Grilled, 100% Antibiotic-Free Chicken Breast, Romaine Lettuce, Cheese, Croutons, Caesar Dressing, Toasted Herb Focaccia in An Organic Wheat Wrap.
Calories	600	900
Monetary Allocation ($)	0	1

<ins>Note</ins>: You will be required to eat at least some of the food selected for implementation during today's session, so consider the options carefully.

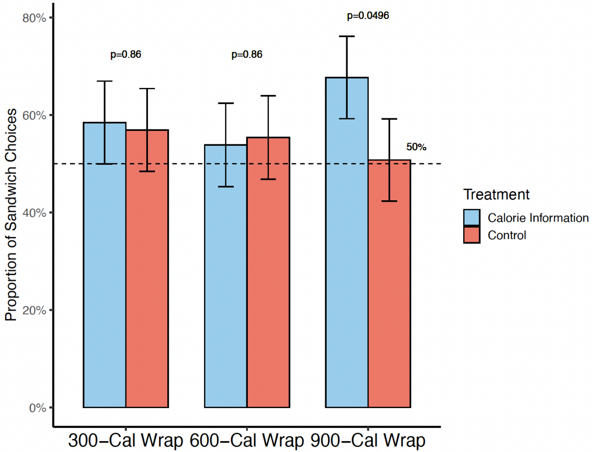
Which Option Do You Prefer?

A

B

Another Example

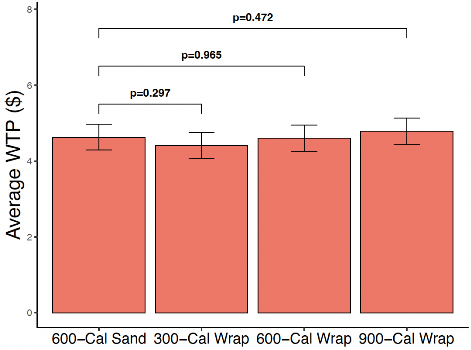
Toney et al. (2023)



- Consistent with the model prediction (50/50 choice) in Control treatment

Why It Works?

Toney et al. (2023)



Why It Works?

Toney et al. (2023)

- Relatively small monetary compensation has a small impact?
- Similar products?
- Is (Food+Money) chosen more than (Food)?
- Coincidence?

Comments on Incentive Compatibility

Halevy et al. (2018) (p.1580)

When subjects understand that the evaluation and main tasks are connected, they may *manipulate* choices in order to maximize their expected choices

1. Make sure that you do not reveal the relation between the two tasks via the instructions and during the experimental procedure
2. It is unlikely that subjects have enough knowledge of the design and successfully manipulate their choices
3. Check what would be expected when subjects combine the two tasks and the fraction of their behavior
   - Halevy et al. (2018) expect the choices to be biased toward EU behavior when subjects combine the tasks (40%)

Conclusion

Is it meaningful to control? Yes!

- Provides a theoretical background
- Suggests insights/directions how to modify models of interest

How to do better?

- Comparative statics
- Interface
- Choose a method
  - Consider the context of the main task, reliability, multiple elicitation
- Adaptive algorithms
  - Structural model

Thank You!

How to Generate Binary Choices in Halevy et al. (2018)

Suppose that $\hat{\theta}_{NLLS} = \{\hat{\beta}_{NLLS}, \hat{\rho}_{NLLS}\}$ and $\hat{\theta}_{MMI} = \{\hat{\beta}_{MMI}, \hat{\rho}_{MMI}\}$
Given a risky portfolio, x^R, compute the CE_i and CE_j where $i, j \in \{NLLS, MMI\}$

Case: $\hat{\beta}_{NLLS}, \hat{\beta}_{MMI} > 0$

Let safe portfolio be $x^s = (CE_i + CE_j)/2$
If $CE_i > CE_j$, then

- $\hat{\theta}_i$ predicts the risky portfolio (x^R)
- $\hat{\theta}_j$ predicts the safe portfolio (x^s)

Experimental Design (Halevy et al. (2018))

Correlation Coefficients in McGranaghan et al. (2024a)

Table A.3: Correlations Between h_{XY} and h'_{XY} by p and r

	(1) $r = 0.1$	(2) $r = 0.2$	(3) $r = 0.3$	(4) $r = 0.5$	(5) $r = 0.8$
			Panel A: $\rho(h_{AB}, h'_{AB})$		
$p = 0.3$	0.256	0.369	0.422	0.372	0.617
$p = 0.5$	0.402	0.464	0.540	0.586	0.696
$p = 0.8$	0.428	0.545	0.395	0.447	0.641
$p = 0.9$	0.314	0.497	0.402	0.519	0.548
			Panel B: $\rho(h_{AB'}, h'_{AB'})$		
$p = 0.3$	0.254	0.492	0.439	0.433	0.545
$p = 0.5$	0.320	0.406	0.445	0.619	0.614
$p = 0.8$	0.564	0.444	0.461	0.475	0.584
$p = 0.9$	0.292	0.514	0.385	0.355	0.483
			Panel C: $\rho(h_{CD}, h'_{CD})$		
$p = 0.3$	0.452	0.453	0.570	0.538	0.541
$p = 0.5$	0.474	0.512	0.410	0.590	0.583
$p = 0.8$	0.435	0.484	0.461	0.389	0.529
$p = 0.9$	0.462	0.431	0.485	0.453	0.432

▸ Discussion

Screenshot of MPL in McGranaghan et al. (2024a)

OPTION A:		OPTION B:
100% CHANCE OF $24	OR	2% CHANCE OF $0 90% CHANCE OF $24 8% CHANCE OF **$24**
100% CHANCE OF $24	OR	2% CHANCE OF $0 90% CHANCE OF $24 8% CHANCE OF **$25**
100% CHANCE OF $24	OR	2% CHANCE OF $0 90% CHANCE OF $24 8% CHANCE OF **$26**
100% CHANCE OF $24	OR	2% CHANCE OF $0 90% CHANCE OF $24 8% CHANCE OF **$27**
100% CHANCE OF $24	OR	2% CHANCE OF $0 90% CHANCE OF $24 8% CHANCE OF **$28**
100% CHANCE OF $24	OR	2% CHANCE OF $0 90% CHANCE OF $24 8% CHANCE OF **$29**
100% CHANCE OF $24	OR	2% CHANCE OF $0 90% CHANCE OF $24 8% CHANCE OF **$30**
100% CHANCE OF $24	OR	2% CHANCE OF $0 90% CHANCE OF $24 8% CHANCE OF **$31**
100% CHANCE OF $24	OR	2% CHANCE OF $0 90% CHANCE OF $24 8% CHANCE OF **$32**
100% CHANCE OF $24	OR	2% CHANCE OF $0 90% CHANCE OF $24 8% CHANCE OF **$33**
100% CHANCE OF $24	OR	2% CHANCE OF $0 90% CHANCE OF $24 8% CHANCE OF **$34**

Discussion

More Algorithms

- Dynamically Optimized Sequential Experimentation (DOSE)
  - (Imai and Camerer, 2018; Chapman et al., 2024)
- Dynamic Experiments for Estimating Preferences (DEEP)
  - (Toubia et al., 2013)
- Adaptive Design Optimization (ADO)
  - (Cavagnaro et al., 2010)
- Sequential Optimal Inference (SOI)
  - (Daviet, 2019; Daviet and Webb, 2023)

How to to better?

References I

Beauchamp, J. P., D. J. Benjamin, D. I. Laibson, and C. F. Chabris (2020). Measuring and controlling for the compromise effect when estimating risk preference parameters. *Experimental Economics 23*, 1069–1099.

Brown, A. L. and P. J. Healy (2018). Separated decisions. *European Economic Review 101*, 20–34.

Cavagnaro, D. R., J. I. Myung, M. A. Pitt, and J. V. Kujala (2010). Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. *Neural computation 22*(4), 887–905.

Chapman, J., E. Snowberg, S. W. Wang, and C. Camerer (2024). Dynamically optimized sequential experimentation (dose) for estimating economic preference parameters. Technical report, National Bureau of Economic Research.

Charness, G., U. Gneezy, and A. Imas (2013). Experimental methods: Eliciting risk preferences. *Journal of economic behavior & organization 87*, 43–51.

References II

Crosetto, P. and A. Filippin (2016). A theoretical and experimental appraisal of four risk elicitation methods. *Experimental Economics 19*, 613–641.

Daviet, R. (2019). Sequential optimal inference for experiments with bayesian particle filters. *Available at SSRN 3320251*.

Daviet, R. and R. Webb (2023). A test of attribute normalization via a double decoy effect. *Journal of Mathematical Psychology 113*, 102741.

Exley, C. L. (2020). Using charity performance metrics as an excuse not to give. *Management Science 66*(2), 553–563.

Exley, C. L. and J. B. Kessler (2024). Motivated errors. *American Economic Review 114*(4), 961–987.

Gillen, B., E. Snowberg, and L. Yariv (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy 127*(4), 1826–1863.

References III

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

Hascher, J., N. Desai, and I. Krajbich (2021). Incentivized and non-incentivized liking ratings outperformwillingness-to-pay in predicting choice. *Judgment and Decision Making 16*(6), 1464–1484.

Im, C. (2024). An experiment on decoy effects for vertically and horizontally differentiated products. Technical report, Working Paper.

Imai, T. and C. F. Camerer (2018). Estimating time preferences from budget set choices using optimal adaptive design. *Unpublished manuscript*.

McGranaghan, C., K. Nielsen, T. O'Donoghue, J. Somerville, and C. D. Sprenger (2024a). Connecting common ratio and common consequence preferences.

McGranaghan, C., K. Nielsen, T. O'Donoghue, J. Somerville, and C. D. Sprenger (2024b). Distinguishing common ratio preferences from common ratio effects using paired valuation tasks. *American Economic Review 114*(2), 307–347.

References IV

Somerville, J. (2022). Range-dependent attribute weighting in consumer choice: An experimental test. *Econometrica 90*(2), 799–830.

Sprenger, C. (2015). An endowment effect for risk: Experimental tests of stochastic reference points. *Journal of Political Economy 123*(6), 1456–1499.

Toney, B., M. A. Palma, M. Segovia, and G. Melo (2023). The instrumental value of calorie information and its effect on food choices. *Available at SSRN 4634107*.

Toubia, O., E. Johnson, T. Evgeniou, and P. Delquié (2013). Dynamic experiments for estimating preferences: An adaptive method of eliciting time and risk parameters. *Management Science 59*(3), 613–640.

Zhou, W. and J. Hey (2018). Context matters. *Experimental economics 21*, 723–756.

Zrill, L. (2024). (non-) parametric recoverability of preferences and choice prediction. *Review of Economics and Statistics 106*(1), 217–229.