Treatment and Welfare Learning for Policymakers

Edward Jee EDJEE@UCHICAGO.EDU
Kenneth C. Griffin Department of Economics
University of Chicago
Chicago, IL 60637, USA

Abstract

I describe a modified Thompson sampling algorithm that jointly learns participant preferences and treatment effects. In contrast to Lin et al. (2022), I design an incentive compatible mechanism to reveal *participant* preferences and not those of the experimenter, even in the presence of strategic behaviour or scepticism of information provided. Finally, by randomising participants into their preferred choice or alternative choices I can directly decompose the average treatment effect (ATE) vectors into an average treatment effect on the treated (ATT) and average treatment effect on the untreated (ATU), which are key parameters of interest for policymakers.

Keywords: Multi-Objective Decision Making, Participant Preferences, Incentive Compatibility, Development Economics, Programme Evaluation, Causal Inference

1. Introduction

Field experiments in development economics often observe multiple, common outcomes across treatment arms and must identify an optimal arm for a policymaker. However, without knowledge of individual's preferences they cannot map changes in outcomes to changes in welfare. Therefore, economists using adaptive trials typically maximise a single outcome or some standardised index. Taking such a decision seriously implies the economist's imposed utility function reflects participants' preferences and these preferences only depend on one feature of the outcome vector. Alternatively, using a standardised index implies participants value the marginal value of a variance weighted good equally across outcomes[1]. Instead, I propose an algorithm that jointly learns participant preferences and treatment arm effectiveness whilst also identifying exactly the policy parameters of interest. Rather than just identifying an average treatment effect (ATE) vector, I can decompose estimates into an average treatment effect on the treated (ATT) and an average treatment effect on the untreated (ATU) and finally, estimate the necessary marginal rates of substitution to aggregate vector valued treatment effects into scalar units of welfare.

My results are most similar to Dewancker et al. (2016); Lin et al. (2022). The latter seeks to learn a decision maker's utility function using pairwise comparisons when a time-consuming or expensive experiment must be run to learn vector valued outcomes. My setting differs in two key ways: First, it is the participants' preferences we wish to learn and not the experimenter's. That is, an experimenter runs the adaptive trial but wishes to learn the preferences of the participants or subjects within the trial. Second, instead of

1. For a static example see Ashraf et al. (2010); Blattman et al. (2017); Bandiera et al. (2017)

simply identifying the ATE, my proposed algorithm identifies both the ATT and the ATU. These parameters are often more relevant to the policymaker's decision to implement the programme considered.

The average treatment effect gives the expected effect of treatment for a randomly selected individual - it averages over the characteristics of individuals within a population. However, one important distinction a policymaker would like to know is the effect of treatment for those that *would choose* the treatment versus those that *wouldn't choose* the treatment if the trial were scaled up to a wider population. In a randomised control trial participants are typically randomised into either treatment or control and have no choice over their treatment status. However when implementing policy, governments can rarely enforce perfect compliance. Therefore, they wish to know the ATT, which measures the effect of treatment for those that choose to takeup the offered policy. On the other hand, the ATU measures the effect of expanding a programme to target a sub-population that previously wouldn't have taken the treatment. Since typically we'd expect individuals to select into treatment based on perceived gains, we'd expect ATT $>$ ATE $>$ ATU.

Moving from experimenter's preferences to trial participants' introduces several complications. A rich literature in economics, pyschology, and experimental fields more generally demonstrate the importance of designing mechanisms such that participants are incentivised to reveal their true preferences (Savage, 1971; Delavande, 2014). When the decision maker and experimenter are one and the same, incentives are clearly aligned and stated preferences can be treated as the ground truth. However, when the experimenter must learn participant preferences he/she must account for the possibility of strategic behaviour and "cheap talk". For instance, participants may misreport their preferences in a bid to increase their expected payoff in response to other participants' choices, or social desirability bias may lead to participants stating preferences to match the experimenter's expectations or social norms (Brownback and Novotny, 2018). Therefore, my work differs from Lin et al. (2022) and methods outlined by (Furnkranz and Hullermeier, 2010) by proposing an incentive compatible preference elicitation step. I incentivise participants to tell the truth by allowing individuals to rank treatment arms and assigning individuals to arms with a probability concordant with their stated preferences. In some sense, my approach "disciplines" participants by ensuring stated preferences have consequences and aligning incentives such that individuals' dominant strategy is to tell the truth.

Furthermore, whilst Lin et al. (2022) treat output vectors from the surrogate model as given and elicit preferences across these vectors, development economists face an environment where information signals are treated with scepticism and both treatment *and* preference experimentation/elicitation in the field is costly. Therefore, I propose a second algorithm with a "structural" model of belief formation that maps posterior signals, which can be imparted by the experimenter cheaply in the field, to participant posterior beliefs.

Finally, this paper seeks to further embed the principles of the Belmont report, published in 1979 after numerous human subject research violations, in randomised control trials run by development economists and other researchers. By enshrining participant preferences at the center of randomised trials, my proposed algorithm speaks directly to respect for persons; beneficence; and justice as outlined by the report. Respect for participants and their preferences, who in development economics often reside in low-income countries, is

particularly important in a field dominated by rich, US-based academics (Stansbury and Schultz, 2022) running trials on those with little agency, income, or human capital.

2. Setup and Method

In this section, I will briefly describe two algorithms. The first assumes a benign enviroment without budget constraints and elicits posteriors directly. The second focuses on a more realistic field environment often faced by development economists in low-income countries where information can be costly to obtain.

The experimenter faces a vector of outcomes $Y_i = \begin{bmatrix} y_i^1 & y_i^2 & ... & y_i^j \end{bmatrix}'$, $Y_i \in \mathbb{R}^j$ for individual i and must determine the optimal treatment arm k with associated J-length reward vector $\boldsymbol{\mu}_k = \begin{bmatrix} \mu_k^1 & \mu_k^2 & ... & \mu_k^j \end{bmatrix}'$, $\boldsymbol{\mu}_k \in \mathbb{R}^j$. Individuals arrive in waves of size N_t. Individuals have knowledge about each treatment arm's expected effectiveness on their own outcomes. Participant's utility functions are parametrised using McFadden (1973)'s discrete choice random utility model[2]: $U_k = V_k + \varepsilon_k$ where U_k corresponds to the utility an individual receives from *choosing* treatment arm k with $V_k = \gamma_1\mu_k^1 + \gamma_2\mu_k^2 + ... + \gamma_j\mu_k^j = \boldsymbol{\mu}_k\boldsymbol{\gamma}'$. Therefore, a participant, if offered the choice of treatment arms, will only choose arm l if $\varepsilon_n < (V_l - V_n) + \varepsilon_l, \forall n \neq l$ where $\varepsilon_k \sim T1EV$, i.e. the Gumbel distribution. Note that γ_j isn't indexed by k - individuals must value a marginal gain in outcome j the same in arm k as arm $l \neq k$.

Algorithm 1 Treatment and participant preference estimation

Generate a prior (Q_0, F_0) over $(\boldsymbol{\mu}_k, \boldsymbol{\gamma})$
for $t = 1, ..., T$ **do**
  Elicit participant posterior beliefs $\boldsymbol{\mu}_k$ using a binarized scoring rule
  Observe participant rankings, K_t and update $F_t(\cdot|K^t, \boldsymbol{\mu}_k(\boldsymbol{\nu}^t))$
  Sample $\boldsymbol{\omega}_t \sim F_t(\cdot|K^t, \boldsymbol{\mu}_k(\boldsymbol{\nu}^t)), \boldsymbol{\nu}_t \sim Q_{t-1}(\cdot|p_k^{t-1}, Y^{t-1})$
  Choose $p_k = \frac{1}{N_t}\sum_{n=1}^{N_t} \mathbb{I}\{\boldsymbol{\mu}_k(\nu_t)\boldsymbol{\gamma}(\omega_t)' > \boldsymbol{\mu}_l(\nu_t)\boldsymbol{\gamma}(\omega_t)'\}, k \neq l$
  Assign participants to treatment arm k with probability p_k using a strategy proof mechanism
  Observe Y_t and update the posterior $Q_t(\cdot|p_k^{t-1}, Y^t)$ over $\boldsymbol{\mu}_k$.
end for

Each wave the experimenter elicits participant posterior beliefs about treatment arm effects using a *binarized scoring rule* (BSR) (Hossain and Okui, 2013). Under a BSR the participant receives a fixed reward when their prediction error is less than some independently generated random number. Since the reward size is fixed the individual is incentivised to report her true beliefs even under a range of risk preferences - only the probability of reward is determined by the realised score and not its size. Hossain and Okui (2013) show that even if a particpant's decision cannot be rationalised by expected utility theory, provided participant preferences satisfy a monotonicity condition the BSR will be incentive compatible and she will report her true beliefs.

2. Any non/semi/fully-parametric choice model could be used here but economists are particularly partial to the multinomial/rank-ordered logit.

Next, the experimenter instructs participants to rank their preferred treatment arms. Again, incentive compatibility is ensured by using a strategy proof mechanism with probability of arm assignment increasing in participant rankings. One example would be the *random serial dictatorship* mechanism whereby participants are randomly ordered from 1 to N_t, assign the first participant their first choice, the next participant their top choice amongst the remaining choices, and so on. Each treatment arm accepts remaining participants until their assignment proportion, p_k[3], is reached. Random serial dictatorship is a commonly used school choice algorithm, partly because of its strategy proofness (Abdulkadirolu and Snmez, 2003). With rankings and participant posteriors in hand the experimenter estimates a rank-ordered logit, updating F_t, where the probability of a given ranking takes the familiar functional form:

$$Pr(r_i|\boldsymbol{\gamma}) = \prod_{k=1}^{K-1} \frac{\exp(V_{ir_{ik}})}{\sum_{l=k}^{K} \exp(V_{ir_{il}})}$$

To calculate assignment probabilities the experimenter draws from the treatment effect, Q_{t-1}, and discrete choice model posterior, F_t, to generate $\boldsymbol{\mu}_k(\nu_t), \boldsymbol{\gamma}(\omega_t)$ draws. Taking the linear combination of these draws, $\boldsymbol{\mu}_k(\nu_t)\boldsymbol{\gamma}(\omega_t)'$, gives posterior arm welfare and p_k is chosen using probability matching in proportion to the probability an arm's welfare is highest. Finally, the experimenter assigns participants to treatment arms, observes Y_t and updates their treatment effect posterior, Q_t.

2.1 Algorithm 2

In many settings eliciting posteriors directly is uneconomical. Enumerators must be trained in the application of BSR and administer the test in the field for each subject (Glennerster, 2017). In contrast, information treatments are cheap and easy to administer via text (Banerjee et al., 2021). Therefore, I propose an alternative algorithm that uses signals, in the form of treatment effect posterior draws, and a structural model of belief formation to estimate marginal rates of substitution across outcomes.

In Algorithm 2 the experimenter chooses a subsample to elicit participants' priors over treatment arm effectiveness[4] and samples N_t draws, *or signals,* from the joint posterior of treatment arm effects. Next, the experimenter individually informs participants of a private signal and asks the individual to rank his/her preferred treatment arms. Estimating a rank-ordered discrete choice model of rankings on signals and normalising estimated coefficients by the first signal coefficient gives the marginal rate of substitution *across signals about outcomes*. Unfortunately, this complicates identification somewhat as we must disentangle how much an individual values an additional unit of an outcome from how sceptical they are about outcome signals. In the interest of brevity I will focus on a conjugate Gaussian updating model with known variance:

3. I describe p_k's calculation shortly. p_k doesn't need to be known before eliciting rankings, only when treatment is assigned.
4. Economists typically ask individuals to allocate beans or stones in intervals to generate belief distributions e.g. see Delavande (2014) for more details.

Algorithm 2 Treatment and structural participant preference estimation

Generate a prior (Q_0, F_0, Π_0) over $\left(\boldsymbol{\mu}_k, \boldsymbol{\gamma}, (\boldsymbol{\mu}_0, \boldsymbol{\tau}_0^{-1})\right)$
for $t = 1, ..., T$ **do**
  if $e_i < \alpha_t, e_i \sim U(0,1)$ **then**
    Elicit participant priors, $\boldsymbol{\mu}_0, \boldsymbol{\tau}_0^{-1}$, using BRS and update Π_t
  end if
  Sample $\boldsymbol{\nu}_t \sim Q_{t-1}(\cdot|p_k^{t-1}, Y^{t-1})$ and inform each participant of a single $\mu_k(\nu_t)$ draw
  Observe participant rankings, K_t, and update $F_t(\cdot|K^t, \boldsymbol{\mu}_k(\boldsymbol{\nu}^t))$ given Π_t
  Sample $\boldsymbol{\omega}_t \sim F_t(\cdot|K^t, \boldsymbol{\mu}_k(\boldsymbol{\nu}^t)), \boldsymbol{\nu}_t \sim Q_{t-1}(\cdot|p_k^{t-1}, Y^{t-1}), \boldsymbol{u} \sim U(0,1)$
  Choose $p_k = \frac{1}{N_t}\sum_{n=1}^{N_t} \mathbb{I}\{\boldsymbol{\mu}_k(\nu_t)\boldsymbol{\gamma}(\omega_t)' > \boldsymbol{\mu}_l(\nu_t)\boldsymbol{\gamma}(\omega_t)'\}, k \neq l$
  Assign participants to treatment arm k with probability p_k using a strategy proof mechanism
  Observe Y_t and update the posterior $Q_t(\cdot|p_k^{t-1}, Y^t)$ over $\boldsymbol{\mu}_k$.
end for

$$
\begin{aligned}
S^i|\mu^i &\sim N\left(\mu^i, \tau^{i,-1}\right) \\
\mu^i &\sim N\left(\mu_0^i, \tau_0^{i,-1}\right) \\
\implies E[\mu'^i|s^i] &= \mu_0^i + (s^i - \mu_0^i)\frac{\tau^i}{\tau_0^i + \tau^i},\ i = 1, ..., j
\end{aligned}
$$

That is, on receiving a signal s for outcome i the participant updates their posterior in proportion to their prior precision and precision of the signal. Defining $\lambda^i = \frac{\tau_0^i}{\tau^i}$ gives:

$$
\frac{ds^i}{ds^l} = \frac{\lambda^i + 1}{\lambda^l + 1}\frac{d\mu^i}{d\mu^l}
$$

and estimated participant prior precision and marginal rates of substitution across signals can be mapped into marginal rates of substitution across outcomes.

Again, the experimenter can now sample from the joint posterior over treatment effects and marginal rates of substitution, construct each arm's posterior welfare, and assign treatment using probability matching in proportion to estimated posterior welfare.

2.2 Policy Relevant Parameters

Finally, the estimated parameter vector $\boldsymbol{\mu}_k$ forms an ATE, typically written as $E[Y(1) - Y(0)]$ using the Fisher-Neyman-Rubin-Quandt model. The ATE describes a policy counterfactual if random individuals were *compelled* to take or not take treatment, D. In reality, policymakers typically cannot mandate treatment in a population and must consider whether to expand or withdraw a programme based on the effects for those who choose, or don't choose, to takeup the offered treatment (Heckman, 2010; Heckman and Vytlacil, 2001).

The parameters of interest in this case corresponds to an ATT $E[Y(1) - Y(0)|D = k]$, and ATU, $E[Y(1) - Y(0)|D = k'], k \neq k'$. By eliciting preference rankings but randomly

assigning individuals to treatment arms I can compare outcomes for those that received their favoured choice and those that don't and subsequently uncover the ATT and ATU.

Simulation Results and Discussion

Table 1 shows results from 100 Monte Carlo draws using 15 rounds of 100 participants per wave with four treatment arms and three outcomes to aggregate across. Simulation parameters are drawn from:

$$\begin{aligned} \boldsymbol{\gamma} &\sim N(\mathbf{0}, I_3) \\ \boldsymbol{\mu}_k &\sim N(\mathbf{0}, I_3),\ k = 1, ..., 4 \\ \eta_i &\sim N(0, 1),\ \varepsilon_{ki} \sim T1EV \end{aligned}$$

where η_i, ε_{ki} represent participant-level outcome and ranking errors respectively. Models are estimated in Stan (Carpenter et al., 2017).

Table 1: Monte-Carlo Results

Assignment Type	Pr(Optimal Arm)	Mean Welfare Rank
Estimated	0.95	1.28
Random Assignment	0.87	3.06
Equal	0.39	2.74
First	0.30	2.92

Assignment type "Estimated" corresponds to Algorithm 1 outlined above and identifies the optimal arm by the end of the trial 95% of the time. In contrast, static random assignment only identifes the optimal arm in 87% of draws. "Equal" corresponds to Thompson sampling maximising a standardised index of the three outcomes whilst "First" only targets the first element of the outcome vector to maximise. Since I use a closed form solution for participant utility, using the multinomial logit and generated $\boldsymbol{\gamma}$ parameters, I calculate average welfare across participants within a simulated draw and rank each algorithm, denoted by "Mean Welfare Rank". As expected, Algorithm 1, which estimates participant preferences directly produces the greatest mean welfare for participants whilst static random assignment the lowest.

In conclusion, by carefulling incentivising research trial participants, through the use of binarized scoring rules and a strategy proof mechanism such as random serial dictatorship, I've shown how to estimate participant preferences and aggregate treatment effects across disparate outcomes into a microfounded estimate of arm arm welfare. When direct posterior elicitation if infeasible or prohibitively expensive, drawing signals from the Thompson posterior over treatment effects is almost exactly the information we wish to impart to participants to generate observale variation in choices.

References

Atila Abdulkadirolu and Tayfun Snmez. School choice: A mechanism design approach. *American Economic Review*, 93(3):729747, 2003. doi: doi:10.1257/000282803322157061. URL https://www.aeaweb.org/articles?id=10.1257/000282803322157061.

Nava Ashraf, Dean Karlan, and Wesley Yin. Female empowerment: Impact of a commitment savings product in the philippines. *World Development*, 38(3):333344, 2010. doi: ISSN0305-750X.doi. URL https://doi.org/10.1016/j.worlddev.2009.05.010.

Oriana Bandiera, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. Labor markets and poverty in village economies*. *The Quarterly Journal of Economics*, 132(2):811870, 2017. doi: doi:10.1093/qje/qjx003. URL https://doi.org/10.1093/qje/qjx003.

Abhijit Banerjee, Arun G. Chandrasekhar, Suresh Dalpath, Esther Duflo, John Floretta, Matthew O. Jackson, Harini Kannan, Francine N. Loza, Anirudh Sankar, Anna Schrimpf, and Maheshwor Shrestha. Selecting the most effective nudge: Evidence from a largescale experiment on immunization. In *Working Paper 28726, National Bureau of Economic Research*. 2021. URL http://www.nber.org/papers/w28726.

Christopher Blattman, Julian C. Jamison, and Margaret Sheridan. Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in liberia. *American Economic Review*, 107(4):11651206, 2017. doi: doi:10.1257/aer.20150503. URL https://www.aeaweb.org/articles?id=10.1257/aer.20150503.

Andy Brownback and Aaron Novotny. Social desirability bias and polling errors in the 2016 presidential election. *Journal of Behavioral and Experimental Economics*, 74:3856, 2018.

Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76(1), 2017.

Adeline Delavande. Probabilistic expectations in developing countries. *Annual Review of Economics*, 6(1):1–20, 2014. doi: 10.1146/annurev-economics-072413-105148. URL https://doi.org/10.1146/annurev-economics-072413-105148.

Ian Dewancker, Michael McCourt, and Samuel Ainsworth. Interactive preference learning of utility functions for multi-objective optimization, 2016. URL https://arxiv.org/abs/1612.04453.

Johannes Furnkranz and Eyke Hullermeier. Preference learning and ranking by pairwise comparison. In *Preference learning*, page 6582. Springer, 2010.

R. Glennerster. Chapter 5 - the practicalities of running randomized evaluations: Partnerships, measurement, ethics, and transparency. In Abhijit Vinayak Banerjee and Esther Duflo, editors, *Handbook of Field Experiments*, volume 1 of *Handbook of Economic Field Experiments*, pages 175–243. North-Holland, 2017. doi: https://doi.org/10.1016/bs.hefe.2016.10.002. URL https://www.sciencedirect.com/science/article/pii/S2214658X16300150.

James J. Heckman. Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature*, 48(2):356–98, June 2010. doi: 10.1257/jel.48.2.356. URL https://www.aeaweb.org/articles?id=10.1257/jel.48.2.356.

James J. Heckman and Edward Vytlacil. Policy-relevant treatment effects. *The American Economic Review*, 91(2):107–111, 2001. ISSN 00028282. URL http://www.jstor.org/stable/2677742.

Tanjim Hossain and Ryo Okui. The Binarized Scoring Rule. *The Review of Economic Studies*, 80(3):984–1001, 02 2013. ISSN 0034-6527. doi: 10.1093/restud/rdt006. URL https://doi.org/10.1093/restud/rdt006.

Zhiyuan Jerry Lin, Raul Astudillo, Peter I. Frazier, and Eytan Bakshy. Preference exploration for efficient bayesian optimization with multiple outcomes, 2022.

Daniel McFadden. Conditional logit analysis of qualitative choice behavior. 1973.

Leonard Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971. doi: 10.1080/01621459.1971.10482346.

Anna Stansbury and Robert Schultz. Socioeconomic diversity of economics phds, 2022. URL https://ssrn.com/abstract=4068831.