Does Personalized Allocation Make Our Experimental Designs More Fair?

Please click here for the most recent version

Bertille Picard*

Aix-Marseille Université, AMSE, CNRS

Preliminary Draft - Please do not circulate

Abstract

Algorithms can optimize treatment allocation within an experimental design. They can progressively identify the most beneficial treatment for the subjects and thus maximize the experiment's overall impact. However, these designs raise concerns for experimentalists and policymakers because they involve transferring decision-making to an algorithm. Are adaptive experiments inherently fairer and thus a preferred choice over traditional randomized controlled trials? In this paper, I propose a comprehensive examination of fairness by considering multiple criteria that can influence the researchers' preference for one design over the other: the possibility to increase the benefits of the experiment for the experimental subjects, the transparency of the decision rule, the absence of discrimination regarding the treatment allocation, the protection of individuals' data. By summarizing and analyzing these distinct criteria through a utility model, I discuss the relative fairness of adaptive experiments and standard randomized controlled trials. Specifically, I show that these different designs align with extreme versions of the fairness utility model, reflecting the pursuit of distinct fairness objectives within experimental settings. I highlight intermediate solutions that can be pursued to reconcile and balance different fairness objectives in experimental designs.

JEL: C46, D31

Keywords: fairness, contextual bandits, randomized controlled trials, machine learning

November 6, 2023

*I would like to thank Guillaume Bied (LISN-CREST), Emmanuel Flachaire (AMSE), Ewen Gallic (AMSE), Morgane Hoffmann (CREST), Pierre Michel (AMSE), Louis Pape (Télécom Paris) and Sarah Vincent (AMSE) for their support and guidance. I am thankful for their time and expertise. I would like to express my gratitude to the participants of the seminars, in particular at AMSE, for their valuable suggestions and feedback. Their input greatly contributed to the development and improvement of this work. Any errors or omissions that would remain are solely my responsibility.

Contents

1 Introduction

Individuals' well-being can be significantly impacted by receiving medical treatment or being subject to a specific public policy. While completely randomized experiments ensure equal chances of receiving treatment, this unconditional equal opportunity alone may not suffice to ensure the acceptability of experiments: as more data accumulates, it may become relevant to consider increasing the number of individuals receiving treatment or, conversely, discontinuing the treatment. Adaptive designs provide promising avenues for the progressive allocation of treatments, addressing ethical concerns and avoiding sub-optimal treatments, especially when there is a possibility to personalize the allocation of treatments. In this paper, I explore the concept of fairness by suggesting that allowing an algorithm to optimize treatment allocations during experiments could be more equitable for experimental subjects. I put forward a tension in experimental contexts between the goals of maximizing individual gains and issues of algorithm fairness, as studied in the machine learning literature.

Indeed, the hope of combining personalization and experimentation exploiting large datasets has become a reality. Nonetheless and beyond what is technically possible, the ethical implications associated with algorithmic decision-making have not yet been fully examined in the field of economic and medical experimentation.

In this paper, I argue that the trade-off betweens different experimental designs can be made based on fairness criteria towards the experimental subjects, but that a preference for one or the other design requires a clear and complex expression of the objectives and fairness criteria addressed. I model the choice of an experimental researcher or evaluation institution with fairness preferences that take into account the interests of experimental subjects. The objective of this work is to rethink the choice of experimental design regarding decision theory and economics of justice.

This model shows that one can interpret the choices of experimental designs as optimization solutions of extreme versions of the fairness objective function. I note that experimental designs lead to a transfer of decision power for treatment allocation from a researcher to an algorithm. A standard randomized controlled trial can be seen as a rudimentary algorithm, consisting of a single instruction to assign individuals to treatment arms. Thus, common concerns about algorithmic decision-making, such as transparency and the use of personal data, may not arise due to the simplicity of the allocation rule, similar to a classical lottery with a single draw per individual. In contrast, these concerns become relevant when the allocation rule evolves during the experiment, where the decision rule is complex and based on the observed outcomes, the shape of the welfare function, and sometimes the data subjects' characteristics, as is the case with adaptive experiments. When allocation probabilities are adjusted based on past data, considerations regarding algorithmic decision-making become more prominent. I also argue that hybrid versions of these designs may be desirable to deal with ethical conflicts in a world where several value systems coexist. In this way, I hope to contribute to a reflection on experimental ethics, value alignment with Artificial Intelligence (AI), and the implementation of just policies.

The causal machine learning literature has recently made considerable progress in personalizing counterfactual estimation (see Knaus et al. 2021 for an overview). At the same time, multi-armed bandits have been studied extensively to maximize the total impact of an experiment, motivated by online learning problems, such as web search and ad placement (see Bubeck and Cesa-Bianchi 2012 for a survey). Recently, a field of research in experimental sciences has been investigating the possibility of importing these models in contexts where evaluation and optimization would be two desirable ends of an experiment (e.g. Caria et al. 2020, Offer-Westort et al. 2021, Athey et al. 2021). By promising to progressively adapt the strategy for allocating treatment arms during the evaluation, these models may appear more acceptable to institutions and experimental subjects, while maintaining the scientific objective of evaluating treatment effects.

At the same time, personalization cannot be achieved without the collection of individual data, which exposes the experimentalist to various criticisms, such as concerns about the sometimes massive collection of data, the lack of transparency, potential algorithmic biases, and other ethical considerations. Recently, these issues have indeed gained increasing significance in the machine learning literature. Within the public debate, numerous applications of machine learning, whether for decision-making or other purposes, have raised concerns. For example, journalists Angwin et al. (2016) accused COMPAS, a tool designed to detect future criminals, of discriminating against black people in the United States. Similarly, Amazon's AI recruitment tool was charged with discriminating against women by journalist Dastin (2018) in Reuters. The concerns related to data collection and algorithmic decision-making are being increasingly documented and discussed across academic fields. In computer science, they have gained significant traction, notably through conferences that specifically address fairness such as the ACM FAccT or AAAIC/ACM conferences. More recently, these concerns have also gained attention in the field of economics, prompting reflection on the place of AI fairness in these two domains (see Cowgill and Tucker 2019). Data collection and algorithmic decision-making are also at the heart of a legal debate that has led to the introduction of regulations such as the General Data Protection Regulation (GDPR) in the EU.

It is important to note that there are many different approaches to fairness and that there is no philosophical consensus on fairness. For example, Wachter et al. (2021) identifies a "critical incompatibility between European notions of discrimination and existing work on algorithmic and automated designs". For practitioners, determining why an experimental design would be more or less fair is therefore a real puzzle. Moreover, preferences of experimental subjects regarding data-driven methods and algorithms versus a human allocation rule are barely documented. The findings on human attitudes towards algorithms are still inconclusive, with some studies suggesting a lack of trust in algorithmic decision-making

(Bigman and Gray 2018, Gogoll and Uhl 2018) while Chugunova and Luhan (2022) suggests a preference for the algorithm, based on a laboratory experiment.

In my paper, fairness refers to a practice or treatment that involves just and equitable treatment without causing harm or discrimination. Formally, I will define fairness in the context of experimental designs as a weighted combination of fairness metrics related to (1) Maximizing short-term outcomes in line with welfare maximization, (2) Simplicity and transparency of the treatment allocation rule, (3) Ensuring equal opportunities to receive the treatment, and (4) Minimizing data collection. I argue that these criteria align with fundamental principles in experimental ethics: experiments should benefit subjects, require informed consent, prevent discrimination, and avoid putting subjects in harm's way[1]. This being said, "What is fair?" is a question that has been the subject of far too many philosophical discussions and confrontations of moral values for an explicit answer to be given. In this discussion, I address common aspects that arise in the literature of economics, philosophy, and (mostly European) law. However, the intention here is not to conclude the debate on this question.

First, I propose a model of the researcher's preferences that would be centered on considerations of fairness to individuals while learning treatment effects. I will make explicit different dimensions of this fairness. Second, I will interpret the objectives of the different experimental designs as solutions of some extreme parameterizations of this preference function. Finally, I will propose an illustration on simulated data.

2 The key elements of the model with multiple criteria of fairness

2.1 Why fairness towards experimental subjects matters to an experimentalist

Collaboration with institutions or policymakers is often necessary for the implementation of controlled evaluations. For them to take place, they must be perceived as acceptable by the subjects of the experiment, since the attitude of data subjects and citizens towards the experimental protocol can significantly influence its implementation. Thus, Dur et al. (2023) show that politicians conform to voters' appreciation of policy experiments to a large extent using representative Dutch survey data. They also show that concerns about the unfairness of the experiments, because for a given period, people and firms are treated unequally, is

[1] Note that the term "fairness" is preferred to the notion of "social justice" in this context for two reasons. First, I will focus on fairness to experimental subjects and the direct impact of the experiment, which excludes a more general discussion of the ultimate distribution of resources among all individuals in a society, which is an important topic in redistributive justice. However, there is a fine line between the two terms and they can sometimes be used synonymously. Second, "fairness" is a term that has recently taken hold in the machine learning literature, and I believe that, without obscuring the history of this term in the field of economics, it carries a strong and close meaning in this second literature and facilitates navigation between these two disciplinary fields.

shared at least to some extent by one-third of the voters. Making experiments fairer through adaptive experiments could lead us to both more and better policies. They provide an answer to a major criticism of randomized controlled trials (RCTs), namely a lack of ethics towards experimental subjects due to the sending of suboptimal treatments to a large number of individuals with the sole aim of establishing clear inferences on treatment effects.

Nevertheless, a dilemma must be solved: the problem of multi-armed bandits has long been discussed in the literature. Multi-armed bandits problem consist in discovering progressively the most successful treatment arms. When individual characteristics are also used to target the treatments, designs are referred to as "contextual multi-armed bandits." As early as 1933, Thompson proposed an algorithm for solving the reward maximization problem in a multi-arm adaptive experimentation setting, called Thompson sampling. These designs have not yet become the gold standard in economics or medicine compared to standard RCTs, which were originally introduced by Neyman in 1923 and Fisher in 1925 (Banerjee et al. 2016). Furthermore, the treatment allocation rule in randomized controlled trials (RCTs) is often considered an ethical argument in itself due to its simplicity and transparency. This is because it provides everyone with an equal chance to receive the treatment. The promise of convergence toward an individual optimal treatment could be a strong ethical argument since the situation of each individual could be improved.

Thus, as a first step, I will identify fairness values that can be put forward in the context of policy or treatment evaluation based on the experimental literature and the economic theory of justice. Some experimental designs neglect certain values, and expliciting these in a single model will therefore allow us to fuel the discussion.

2.2 General Framework

Firstly, let us propose a very simple framework. The subjects of the experiment correspond to a population of individuals numbered by an index i taking integer values between 1 and N. To simplify the discussion, I consider a problem with only two treatment arms, for example, a control group (0) and a treatment group (1). Hereafter, I will refer to group (0) as the "untreated" group, and group (1) as the "treated" group, but the reasoning applies to the case where group 0 receives a specific treatment denoted A and group 1 receives a specific treatment denoted B. The researcher's job is to decide $p_1, ...p_n$, the probabilities of allocating treatment 1 to each individual, while maximizing a fairness program over the individuals. The researcher's utility will thus depend on the value given to different justice criteria that I propose to develop in the following subsections. Following the framework of potential outcomes (Rubin 1974), I consider that by being treated, an individual receives an outcome $Y(1)$, and by not being treated receives $Y(0)$. The variable indicating the fact that the individual is treated denoted T, takes the value 0 (control) or 1 (treated). The treatment can positively or negatively affect the individual's outcome. Furthermore, the individual will be described by two binary characteristics:

- Variable G is 1 if the individual belongs to a protected group and 0 otherwise. G can represent a group according to which it is not considered acceptable to discriminate, such as gender, race, disability status, sexuality, religion, etc.
- Variable X which is another binary characteristic according to which it is acceptable to discriminate.

X and G are potentially correlated, and (X, G) can determine $Y(1)$ and $Y(0)$.

2.3 Fairness considerations in experimental setups

I identify here 5 main types of fairness values that could be put forward in assessing the ethics of one experimental design versus another: (1) improvement of everyone's situation, and in particular of the worse-off, (2) simplicity and transparency of the treatment allocation rule, (3) equality of opportunity, i.e. permitting differentiation of treatment allocation chances solely based on specific, clearly defined variables that are deemed acceptable for differentiation, and (4) protection of individual data, that is the minimization of personal data collection.

This paper has interesting similarities with Viviano and Bradic (2022)'s work on fair policy targeting. However, the fact that the problem examined here concerns treatment allocation in an experimental context and not a reflection on welfare programs from the policymaker's point of view makes an important difference. In Viviano and Bradic (2022)'s paper, several definitions of fairness are also proposed in the context of policy targeting, some of which echo the dimensions discussed here. The agent can indeed choose a definition of unfairness (1) based on observed differences in policy allocation, (2) based on welfare disparities, (3) based on an individual's incentive to reveal a sensitive attribute and (4) based on notions of counterfactual fairness. The approach of this paper is, in my understanding, very interesting in a perspective where we would like to guide a policymaker who would like to target policy on a population and would have preferences for a type of fairness. In this paper, I focus more on experimental considerations, resulting in slightly different choices of fairness dimensions. The definition based on observed differences in allocation can be linked to our first criterion of transparency of the allocation rule. Indeed, by choosing a probability π of allocating the treatment before running the experiment in a standard RCT, the experimentalist ensures that the unfairness criterion based on observed differences in allocation is null. The ethical role of my criterion is simply stronger in this paper: it is not just a question of ensuring equal treatment between individuals, but also of ensuring transparency of the rule by specifying the probability upstream, to be able to announce it to the experimental subjects and allow informed consent. Moreover, in an experimental context, the decision to collect sensitive data is taken by the experimentalist, and formulation in the form of an incentive may not be appropriate in this context, which is why I opt for modeling in terms of costs. I consider that during the experimental phase, the experimentalist struggles to define measurable welfare, but I take into account the possibility that a short- or medium-term outcome may be observable and desirable. Rather than considering inequalities in long-term welfare, I consider inequalities in access to short-term treatment, in the absence of knowledge of the effects on long-term outcomes.

It should be noted that in practice, other criteria of comparison concerning ethics in a less direct way could also be examined, such as the cost of setting up the experiment, the value attributed to the scientific result for the community as a whole, and the externalities resulting from the dissemination of knowledge derived from the experiment. These mechanisms may vary from one experimentation to another, and depend on the organization of the research and the usefulness of the individuals not participating in the experimentation. Their integration could be left to future work.

2.4 Criterion 1: maximizing a short-term outcome improving welfare

In 1979, the Belmont Report was finalized, laying down 3 principles designed to protect the rights of research subjects: (1) respect for persons, (2) beneficence, and (3) justice. These principles have served as a significant inspiration for promoting ethical conduct in experimental research, especially for clinical trials. To be consistent with the second principle, that of beneficence, we can consider that there are two rules to follow: (1) do no harm and (2) increase potential benefits and decrease possible adverse events or harms (Miracle 2016). This second rule in particular could be an argument in favor of adaptive experiments, and even more so of contextual adaptive experiments, that personalized the allocation based on the observed characteristics. The proposed algorithms explicitly seek a solution to maximize benefits.

When the outcome $Y(1)$ or $Y(0)$ is an outcome that increases the individual's satisfaction, a reasonable goal is to maximize the outcome $Y^{\text{obs}} = Y(1)T + Y(0)(1-T)$. In other words, it may be desirable to send the individual to the treatment that maximizes the impact on a variable universally considered desirable.

2.4.1 The virtues of the contextual bandit goal from a utilitarian perspective

Let us consider a restrictive but relevant case where the objective of the contextual bandit is aligned with a social welfare function. In a contextual bandit, it is possible to allocate the optimal treatments according to the characteristics, thus personalizing the allocation of the treatment arms. The objective is thus a personalized allocation:

$$\max_{t \in \{0,1\}^N} \sum_{i=1}^{N} Y_i(t_i) \mathbb{1}\left(T_i = t_i\right) \tag{1}$$

Let us consider some simple but strong conditions justifying the social desirability of the solution targeted by contextual bandits. Suppose a researcher has a preference relation R to rank the different treatment allocation options. Let us denote A the set of possible allocations of the treatment to N individuals. A contains 2^N elements, corresponding to the outcomes of N random draws among $\{0,1\}$. Let us note $U_1(\cdot),...U_N(\cdot)$ the individual utilities associated with an allocation a. Let $A_i \subset A$ be the set of allocations such that i receives the treatment, and $\bar{A}_i \subset A$ the set of allocations such that i does not receive the treatment. Let us assume that individual utility is purely selfish, i.e. that an experimental subject is only concerned with his own allocation[2]:

Assumption 1: individualism

$$\forall i \in [\![1,N]\!], \begin{cases} \forall\,(a_0,a_1) \subset A_i^2, & U_i\,(a_0) = U_i\,(a_1) \\ \forall\,(a_0,a_1) \subset \bar{A}_i^2, & U_i\,(a_0) = U_i\,(a_1) \end{cases}$$

Let us define W_i as the unit vector where every component is 0 except the i^{th} element that is a 1. In other words, we have $A_i = \{a \in A \mid \langle W_i, a\rangle = 1\}$ and $\bar{A}_i = \{a \in A \mid \langle W_i, a\rangle = 0\}$. We can state a second assumption that will systematically make the solution of the contextual bandit desirable:

Assumption 2: unconditional desirability of Y

$$\forall i \in [\![1,N]\!], \quad Y_i\,(\langle W_i, a_0\rangle) \geq Y_i\,(\langle W_i, a_1\rangle) \Leftrightarrow U_i\,(a_0) \geq U_i\,(a_1)$$

According to this assumption, an individual always prefers that the treatment increases their outcome[3].

Under these assumptions, it then becomes clear that the contextual bandit's goal becomes consistent with the utilitarian criterion:

$$\forall (a_0,a_1) \subset A^2,\, a_0 \text{ R } a_1 \Leftrightarrow \sum_{i=1}^{N} U_i(a_0) \geq \sum_{i=1}^{N} U_i(a_1)$$

In this restrictive case, letting an algorithm converge toward the solution of its objective seems desirable. The targeted solution will guarantee an improvement in everyone's situation. The strong Pareto principle will be guaranteed, as will the principle of Rawls' difference, since by aiming at the improvement of the situation of each person it allows the improvement of the situation of the most disadvantaged. They would not be guaranteed in a standard RCT. This criterion is also compatible with the principle of anonymity, implying the principle of non-

[2] This assumption implies that there are no peer effects.

[3] In the case where $Y(\cdot)$ is unconditionally non-desirable, like the number of days staying sick, one can state that $-Y(\cdot)$ will be unconditionally desirable.

dictatorship since the utilities in the above sums can easily be inverted without changing the value of the sum. It would then be detrimental not to algorithmically search for the optimal solution during experimentation. A particular case where these assumptions are credible is perhaps that of health. If the experimental subjects are indifferent to each other, and if everyone wants to improve their own health, then these strong assumptions hold.

2.4.2 The problem with non-contextual bandits

This restrictive case also helps us to understand why customization is very attractive, beyond the general increase of impact. Let us suppose that, instead of looking for a personalized solution through a contextual bandit, we look for an optimal solution on average without taking the context into account. The set of alternative allocations A is then restricted to two options: $(0, ...0)$, i.e. allocate the treatment to nobody, and $(1, ...1)$, i.e. allocate the treatment to everybody. The objective of the algorithm is then :

$$\max_{t \in \{0,1\}} \sum_{i=1}^{N} Y_i(t) \mathbb{1} \left(T_i = t \right) \tag{2}$$

In this case, the objective is no longer aligned with the social preferences described above. In particular, there is a risk of violating the principle of non-dictatorship. Indeed, if a subgroup sees its situation deteriorate while the majority of experimental subjects have a strong positive treatment effect, the algorithm may choose to attribute the treatment to all. In this case, exploiting a large amount of individual information to personalize the targetting of the treatment appears more desirable.

2.4.3 The overall impact translated as a fairness criterion

In the model, I propose a fairness criterion closely related to the program (1), except that I will reason in terms of expected utility:

$$\max_{p \in [0,1]^N} \sum_{i=1}^{N} \big(p_i \cdot Y_i(1) + (1 - p_i) \cdot Y_i(0) \big) \tag{3}$$

Reasoning in terms of expected utility will be helpful to highlight intermediate solutions where we would prefer to give the individually optimal allocation while limiting the differences between the weights for different categories of individuals.

Although there are clear arguments in favor of using an algorithm that maximizes a certain outcome variable from the point of view of fairness towards the subjects of the experiment, the assumptions set out above remain debatable. Firstly, one can assume that there are cases where individuals may be reluctant to see Y positively affected by their neighbors if inequalities increase between individuals. This could be the case if a public policy had a stronger effect on the income or resources of the better-off, and increased inequalities between test subjects. For example, a policy that increases the test scores of every student but that increases the gap between students from less and more privileged backgrounds could be viewed as non-desirable. Moreover, the desirability of outcome Y must be studied on a case-by-case basis. For example, if Y represents a return to employment or its proxy, this may only be desirable on condition that the new job is stable. Finally, in practice, personalizing the treatment during the experiment presupposes that the algorithm can discover the best treatments for each sub-group of the population. A sub-group that is poorly represented in the dataset could suffer from poorer algorithm performance, and the principle of non-dictatorship would then be neglected. This is why it is also necessary to study the arguments against personalization.

2.5 Criterion 2: simplicity and transparency of the treatment allocation rule

The uniform allocation rule is giving each individual the same probability of receiving one treatment arm over another. I will refer to it as the "simple and transparent allocation rule." This treatment allocation rule is probably the most cognitively accessible rule for the researcher and the experimental subject. It amounts to saying that each individual has a probability of say $\pi \in [0, 1]$ receiving the treatment. It is also transparent in the sense that it is known in advance of the experiment, and does not depend on an algorithm, or the impact of the treatment. If an individual asks specifically what is his or her chances of being treated, the uniform rule is natural. In the case of a two-treatment problem, it corresponds to the parameter of a Bernoulli distribution and can be illustrated by the problem of flipping a coin that is possibly unbalanced. In addition to its simplicity, it ensures that the experimental subject understands the conditions of the experiment. The rule of uniform allocation of the treatment also respects a principle of parity, since, whatever the subgroup to which the individual belongs, the chances of being treated are uniformly distributed, regardless of their subgroup. In the field of responsible AI, statistical parity consists of an equitable distribution of AI outcomes between groups. An experimental design involves selecting treatment allocation probabilities for each individual. In the case of a standard Randomized Controlled Trial (RCT), the allocation probabilities are predetermined by the experimentalist, will be equal to $\pi \in [0, 1]$, and are ensured to be respected by design.

Thus, in the model, this preference for this simple rule is reflected in an absolute distaste for any deviation of the treatment allocation probabilities from π. In other words, this fairness criterion is met if and only if:

$$\prod_{i=1}^{N} \mathbb{1}\big(p\left(x_i, g_i\right) = \pi\big) = 1$$

2.6 Criterion 3: equal opportunity to receive the treatment

If X is the only variable determining the probability of treatment allocation, then :

$$\forall (X,G), \quad p(X,G) = E[T \mid X, G] = E[T \mid X]$$

A simple inequality observed between the allocation probabilities of the $G = 1$ and $G = 0$ groups is therefore not sufficient. One way of modeling distaste for inequality of opportunity could be to penalize the part of this difference not explained by X. In other words, one interpretation might be to decompose this difference using the Kitagawa-Oaxaca-Blinder method (Kitagawa 1955, Oaxaca 1973, Blinder 1973). This decomposition method is widely used in the inequality literature to explain the gender wage gap or inequalities between two countries for example. When the average allocation weights in the $G = 0$ and $G = 1$ groups are different, denoting $p(X, g)$ the random variable $\mathbb{E}[T \mid X, G = g]$, we observe :

$$\mathbb{E}[p(X,G) \mid G = 1] - \mathbb{E}[p(X,g) \mid G = 0] \neq 0$$

The question then arises as to whether this observed difference stems solely from differentiation according to X or from discrimination according to G for equal X characteristics. The average difference in observed probabilities can be broken down into a part explained by differences in X and an unexplained part. Let us denote $p(X, r)$ the reference probability chosen for this decomposition (see Flachaire and Picard 2023). The difference in probabilities can be decomposed as follows:

$$\begin{aligned}
\underbrace{\mathbb{E}[p(X,G) \mid G=1] - \mathbb{E}[p(X,G) \mid G=0]}_{\text{Observed difference in probabilities of being treated}} &= \underbrace{\mathbb{E}[p(X,G) \mid G=1] - \mathbb{E}[p(X,r) \mid G=1]}_{\text{Unexplained advantadge for group 1}} \\
&- \underbrace{\mathbb{E}[p(X,G) \mid G=0] - \mathbb{E}[p(X,r) \mid G=0]}_{\text{Unexplained disadvantage for group 0}} \\
&+ \underbrace{\mathbb{E}[p(X,r) \mid G=1] - \mathbb{E}[p(X,r) \mid G=0]}_{\text{Part due to differences in characteristic } X}
\end{aligned}$$

The objective is to minimize the unexplained parts of the observed difference. As a reference here, I propose to refer to the Neumark's reference group (Neumark 1988). The part to be minimized to take account of this fairness criterion based on a decomposition in means is therefore:

$$\sum_{i=1}^{N} \left(\left(p\left(x_i, g_i\right) - \mathbb{E}\left[T \mid X = x_i\right] \right) \left(\frac{g_i}{\#\left\{i \mid g_i = 1\right\}} - \frac{(1 - g_i)}{\#\left\{i \mid g_i = 0\right)} \right) \right)^2$$

where the negative and positive deviations are treated equivalently due to the square applied to any of them. Note that in practice, the choice of reference group is important for estimating this unexplained part. Alternatives to this choice may be considered as discussed in Flachaire and Picard (2023).

It should be noted that one objective could also have been to minimize inequalities about an outcome affected by the intervention, which we would like to see fairly distributed among agents. However, I think that decomposing an outcome rather than the probability of receiving the treatment has its drawbacks. First of all, we need to ensure that there is a genuine preference for a fair distribution of this outcome. On the other hand, interventions may aim to have an impact on several outcomes, which may involve knowing how to combine these outcomes and which inequality to reduce. Finally, impacts on outcomes are generally unknown, and medium-term outcomes may be difficult to estimate during the experiment. This is why I prefer to decompose inequality of access to treatment.

2.7 Criterion 4: data minimization (Data Protection)

Using personal data to target an intervention may be costly in terms of fairness, even if it seems acceptable to use the characteristics to train the decision algorithm. Data regulations such as the European GDPR advocate minimizing the collection of data according to the purpose of the data processing. The principle of minimization states that personal data should be adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed. Indeed, the collection of multiple personal data favors the identification of individuals in the database and thus exposes the individual to the risk of privacy violation, even when observations are pseudonymized with an identifier or number. This cost can be seen as internalized by the researcher as a moral cost or a monetary cost considering the exposition of legal sanctions. It is considered here as fully exogenous and independent of the context. I will introduce this concern as a cost borne for data collection. This means that, to unlock the possibilities of personalization of the allocation, the agent in the model will pay a cost c_x to obtain the right to use X and c_g to obtain the right to use G as determinants of the probabilities for each individual[4].

2.8 Necessary condition: learning and inference

In many cases, the primary objective of the experimentalist is to infer the average treatment effect (ATE) or the conditional average treatment effects (CATE), even before considering the fairness of the experiment. Making reliable inferences, even on sub-groups of the population, is not a fairness objective in itself, even if in some cases, ignoring the effects of treatment on sub-populations can be perceived as an ethical issue. For example, in the medical field, a lack of knowledge of treatment effects for pregnant women can be seen as a research ethics issue. In most cases, however, I consider the possibility of making inferences to be more of a “selfish” concern for the researcher, hence my reluctance to list this directly as a fairness

[4]The observation of Y can be considered as costly too from this point of view, but since the outcome will in any practical case be observed for experimental purposes, I do not incorporate it in my model.

objective. However, it is a crucial issue in experimentation and the very motivation for setting up an experiment.

Moreover, this dimension is important for this discussion as it may lead to a preference for one design over the other. Standard RCTs are a gold standard because they enable simple and reliable inference, at least for the population included in the experiment. Inference in adaptive experiments is more complicated in practice. I will assume here that the framework for inference is frequentist. Indeed, although the Bayesian approach may seem natural in this context of progressive data collection, and in particular when allocation strategies are based on a Bayesian model as is the case for Thompson sampling, it is not the most common approach in the field of public policy and medical treatment evaluation. Furthermore, purely Bayesian frameworks may be more difficult to adapt to contexts of intensive exploration of fine heterogeneity, for example with machine learning techniques in the context of high-dimensional heterogeneity[5], which undermines the hope of personalizing treatment and avoiding problems of group dictatorship (see discussion in section 2.4.2).

Adaptive experiments have a feature that brings them closer to standard RCTs: the allocation weights remain random for each individual. What's more, as these allocation weights are purely decided by the algorithm, they are traceable and known, so there's no need to re-estimate them, and we know the variables that determine them[6]. Thus, it is no less possible to conduct unbiased inference in this context[7] by reweighting. In this way, the mean treatment effect and conditional effects can be estimated using the Inverse Probability Weighting (IPW) estimator:

$$
\begin{aligned}
CATE(x,g) &= \mathbb{E}\big[Y(1) - Y(0) \mid X = x, G = g\big] \\
&= \mathbb{E}\left[Y^{\text{obs}} \frac{\mathbb{1}\{T=1\}}{p(x,g)} - Y^{\text{obs}} \frac{\mathbb{1}\{T=0\}}{(1-p(x,g))} \mid X = x, G = g\right]
\end{aligned}
$$

These conditional mean effects can be calculated from the empirical counterpart :

$$
\widehat{CATE}(x,g) = \frac{1}{n_{x,g}} \sum_{i=1}^{n_{x,g}} \left[Y_i^{\text{obs}} \frac{\mathbb{1}\{T_i=1\}}{p(x,g)} - Y_i^{\text{obs}} \frac{\mathbb{1}\{T_i=0\}}{(1-p(x,g))}\right]
$$

Where $\sum_i^{n_{x,g}}$ is a sum over all individuals satisfying $X = x$ and $G = g$. This estimator is unbiased and easy to compute since the draw probabilities are perfectly known by de-

[5]One could, however, discuss this statement by mentioning work carried out in a Bayesian framework and including a regularization parameter, and mention for example Chapelle and Li (2011).

[6]In short, the fact that drawing weights vary from one individual to another is no more a problem for constructing unbiased estimators than it is in the context of stratified experiments.

[7]We refer the reader, however, to Zhan et al. (2021) for a discussion of the need to correct estimates to ensure asymptotic normality in a context where time dependence prevents observations from being considered independent and identically distributed.

sign, even in an adaptive experiment. Its known drawback, however, is its sensitivity when the denominator probabilities take values very close to 0 and 1. Now, in the context of an adaptive experiment, a major risk is that of eventually converging toward extreme draw probabilities. The quality of the inference may then depend on whether we manage to maintain draw probabilities bounded away from 0 and 1. In the literature on treatment effects, a common practice is to trim observations with draw probabilities that are too extreme, i.e. to remove them from the estimation. Trimming thresholds often used are, for example, 0.01 and 0.99, or even 0.05 and 0.95. Trimming can be seen as a choice to favor the internal validity of estimates over external validity (see Imbens and Rubin 2015, chapter 16), i.e. to obtain results that are less generalizable over the whole population, but more reliable for the trimmed sample. That said, in a controlled environment, an experimentalist can avoid making this sacrifice by not assigning extreme draw probabilities by design. I consider that the choices of treatment allocation weights are made under the constraint:

$$\forall (x,g) \in \{0,1\}^2, \quad \alpha \leq p(x,g) \leq 1-\alpha$$

where $\alpha \in [0, 0.5]$ is a parameter that has to be chosen by the experimentalist and that would correspond to her preferred trimming threshold. This constraint is designed to avoid the need for trimming and to promote the validity of the inference over the whole sample.

3 Theoretical model

3.1 General utility model

Based on the fairness considerations above, I built the following program for a researcher who would like to base the assignment probabilities of the treatment on the different individuals:

$$\begin{aligned}
\max_{(p(x_i,g_i))_{i=1}^N} \quad & \omega_1 \prod_{i=1}^{N} \mathbb{1}\left(p\left(x_i, g_i\right) = \pi\right) \\
& -\omega_2 \sum_{i=1}^{N} \left(\left(p\left(x_i, g_i\right) - \mathbb{E}\left[T \mid X = x_i\right]\right) \left(\frac{g_i}{\#\left\{i \mid g_i = 1\right\}} - \frac{(1-g_i)}{\#\left\{i \mid g_i = 0\right)} \right) \right)^2 \\
& +\omega_3 \left[\sum_{i=1}^{N} \left(p\left(x_i, g_i\right) \mathbb{E}\left[Y_i(1) \mid X = x_i, G = g_i\right] + \left(1 - p\left(x_i, g_i\right)\right) \mathbb{E}\left[Y_i(0) \mid X = x_i, G = g_i\right] \; - c_x - c_g\right) \right] \\
\text{such that} \quad & \forall i \in [\![1, N]\!], \alpha \leq p(x_i, g_i) \leq 1 - \alpha
\end{aligned}$$

In this model, ω_1 is greater than 0 when there is a concern for a simple assignment rule. This goal is not incompatible with a welfare maximization program, because there can be cases where the optimal assignment rule in terms of cumulative individual gains is a uniform rule (for example, giving everyone the treatment). When there is a dislike for discrimination against the sensitive variable G, then ω_2 should be strictly greater than 0. Note that I consider that the dislike goes here in both directions, meaning in particular that there is both a dislike for positive discrimination as well as discrimination against the privileged group. When one goal is to maximize the gains by personalizing the treatment allocation, then ω_3 is greater than 0. This personalization may have a positive cost c_x or c_g given the fourth consideration of data minimization. The optimization problem above may be expressed also referring to the conditional treatment effects $\tau(x_i, g_i) = \mathbb{E}[Y_i(1) \mid X = x_i, G = g_i] - \mathbb{E}[Y_i(0) \mid X = x_i, G = g_i]$ and a constant $c_i = \mathbb{E}[Y_i(0) \mid X = x_i, G = g_i] - c_x - c_g$ as follows:

$$
\begin{aligned}
\max_{(p(x_i,g_i))_{i=1}^N} \; & \omega_1 \prod_{i=1}^{N} \mathbb{1}\left(p\left(x_i, g_i\right) = \pi\right) \\
& -\omega_2 \sum_{i=1}^{N} \left(\left(p\left(x_i, g_i\right) - \mathbb{E}\left[T \mid X = x_i\right]\right) \left(\frac{g_i}{\#\left\{i \mid g_i = 1\right\}} - \frac{(1-g_i)}{\#\left\{i \mid g_i = 0\right)} \right) \right)^2 \\
& +\omega_3 \left[\sum_{i=1}^{N} \left(p\left(x_i, g_i\right) \tau(x_i, g_i) + c_i\right). \right] \\
\text{such that} \quad & \forall i \in [\![1, N]\!], \quad \alpha \leq p(x_i, g_i) \leq 1 - \alpha
\end{aligned}
$$

This reformulation allows us to see what is really important in the problem: the treatment effects themselves, rather than the potential outcomes.

3.2 Insights from a simplified framework

3.2.1 Simplified Model Version

In concrete applications, optimal probabilities are the same for a given set of characteristics (x, g), meaning that for all $i \in [\![1, N]\!]$, $(k, l) \in \{0, 1\}^2$, I let $p_{kl} := p\left(x_i = k, g_i = l\right)$[8]. The search for optimal weights simplifies to the search for 4 probabilities. Let us similarly denote $\tau_{kl} = \tau(k, l)$.

For simplicity, I will assume here that the population contains exactly one-quarter of individuals of each type, i.e. a uniform distribution within the population of the characteristics $(G, X) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, we can simplify the above model for illustration purposes. Indeed, each individual being described by x_i and g_i, and each type being equiprobable, we have:

[8]In some cases actual weights can be slightly different from one observation to another for exploration purpose but they are drawn from a distribution centered on the same value.

$$
\begin{aligned}
\mathbb{E}\left[T \mid X = x_i\right] &= \mathbb{E}\left[T \mid X = x_i, G = 1\right]\mathbb{P}\left(G = 1 \mid X = x_i\right) \\
&\quad + \mathbb{E}\left[T \mid X = x_i, G = 0\right]\mathbb{P}\left(G = 0 \mid X = x_i\right) \\
&= p\left(x_i, 1\right) \times \frac{1}{2} + p\left(x_i, 0\right) \times \frac{1}{2} \\
&= \frac{p\left(x_i, 1\right) + p\left(x_i, 0\right)}{2} \\
&= \frac{p_{x_i 1} + p_{x_i 0}}{2}
\end{aligned}
$$

Moreover, solving the general program amounts in this case to solving a simplified program for each of the 4 types of individuals, each having a potential outcome when treated ($T = 1$) and not treated ($T = 0$). Finally, the overall goal simplifies to:

$$
\begin{aligned}
\max_{(p_{00},\, p_{01},\, p_{10},\, p_{11}) \in [0,1]^4} \quad & \omega_1 \mathbb{1}\left\{p_{00} = \pi\right\} \mathbb{1}\left\{p_{01} = \pi\right\} \mathbb{1}\left\{p_{10} = \pi\right\} \mathbb{1}\left\{p_{11} = \pi\right\} \\
& - \frac{\omega_2}{8}\left(\left(p_{00} - p_{01}\right)^2 + \left(p_{10} - p_{11}\right)^2\right) \\
& + \omega_3 \sum_{(k,l) \in \{0,1\}^2} \left(p_{kl}\tau_{kl} + c_{kl}\right)\Big) \\
\text{such that} \quad & \forall (x, g) \in \{0,1\}^2, \quad \alpha \leq p(x, g) \leq 1 - \alpha
\end{aligned}
$$

where $\forall (k, l) \in [0, 1]^2$, $c_{kl} = \mathbb{E}\left[Y(0) \mid X = k, G = l\right] - c_x - c_g$. Thus, we can see that optimal solutions will only depend on π, ω_1, ω_2, ω_3, τ_{00}, τ_{01}, τ_{10} and τ_{11}. Knowledge of preferences for the three dimensions of fairness, as well as knowledge of treatment effects, is therefore sufficient to determine optimal allocations.

3.2.2 Solving the program

Starting by discarding solutions of the type $p^\star = (\pi, \pi, \pi, \pi)$ (to reintroduce them in the end), the objective function takes the form of a sum of two sub-functions that can be optimized separately: a function of (p_{00}, p_{01}) and a second function of (p_{10}, p_{11}). In other words, probability choices are made in pairs sharing the same x, and we can anticipate the (p_{00}, p_{01}) allocations without the need for any further information. The two optimization problems being quite similar, let us focus on the function of (p_{00}, p_{01}) to understand the optimal allocations:

$$
f(p_{00}, p_{01}) = -\frac{\omega_2}{8}\left(p_{00} - p_{01}\right)^2 + \omega_3\left(p_{00}\tau_{00} + p_{01}\tau_{01}\right)
$$

Let $D = \left\{(p_{00}, p_{01}) \in \mathbb{R}^2 \mid p_{00} - \alpha \geqslant 0, 1 - \alpha - p_{00} \geqslant 0, p_{01} - \alpha \geqslant 0, 1 - \alpha - p_{01} \geqslant 0\right\}$. Function f being defined on the compact D, it reaches a maximum. The maximum can be reached for either values in $]\alpha, 1 - \alpha[\times]\alpha, 1 - \alpha[$ or on the borders. When reaching a maximum for values at the interior, the candidates verified that the gradient is zero. Otherwise, values at the border should be considered.

Interestingly, there are few cases with interior solutions. Interior solutions arise in two cases: first, when neglecting the last dimension of fairness, that is, assuming $\omega_3 = 0$. In this case, the agent only cares for equality of opportunity concerns and can decide any $(p^\star_{00}, p^\star_{01})$ in the square $[\alpha, 1-\alpha] \times [\alpha, 1-\alpha]$ such that $p^\star_{00} = p^\star_{01}$. In a less radical framework though, interior solutions arise whenever $\tau_{00} = \tau_{01}$ are perfectly symmetrical[9], provided that the treatment effects are not too extreme ($\mid 4 \cdot \omega_3 \cdot \tau_{00}/\omega_2 \mid \leq 1 - 2\alpha$). In these cases, the optimal allocation verify $\mid p^\star_{00} - p^\star_{01} \mid = \mid 4 \cdot \omega_3 \cdot \tau_{00}/\omega_2 \mid$, the lowest probability being of course $p^\star_{00}$ if $\tau_{00} \leq 0$ and $p^\star_{01}$ otherwise.

In every other cases, solutions have to be found on the border. Depending on the relationship between each treatment effect and the initial preferences ω_2 and ω_3, one can compute the points yielding the highest value of $f(\cdot, \cdot)$ on each border. The 4 points can be obtained using table 1 (see proof in appendix):

	$\tau_{00} \leq \frac{\omega_2}{4\omega_3}(2\alpha - 1)$	$\frac{\omega_2}{4\omega_3}(2\alpha - 1) < \tau_{00} \leq 0$	$0 < \tau_{00} \leq \frac{\omega_2}{4\omega_3}(1 - 2\alpha)$	$\frac{\omega_2}{4\omega_3}(1 - 2\alpha) < \tau_{00}$
$\tau_{01} \leq \frac{\omega_2}{4\omega_3}(2\alpha - 1)$	(α, α) $(1-\alpha, \alpha)$ (α, α) $(\alpha, 1-\alpha)$	(α, α) $(1-\alpha, \alpha)$ (α, α) $\left(1-\alpha+\frac{4\omega_3}{\omega_2}\tau_{00}, 1-\alpha\right)$	(α, α) $(1-\alpha, \alpha)$ $\left(\frac{4\omega_3}{\omega_2}\tau_{00}-\alpha, \alpha\right)$ $(1-\alpha, 1-\alpha)$	(α, α) $(1-\alpha, \alpha)$ $(1-\alpha, \alpha)$ $(1-\alpha, 1-\alpha)$
$\frac{\omega_2}{4\omega_3}(2\alpha - 1) < \tau_{01} \leq 0$	(α, α) $\left(1-\alpha, 1-\alpha+\frac{4\omega_3}{\omega_2}\tau_{01}\right)$ (α, α) $(\alpha, 1-\alpha)$	(α, α) $\left(1-\alpha, 1-\alpha+\frac{4\omega_3}{\omega_2}\tau_{01}\right)$ (α, α) $\left(1-\alpha+\frac{4\omega_3}{\omega_2}\tau_{00}, 1-\alpha\right)$	(α, α) $\left(1-\alpha, 1-\alpha+\frac{4\omega_3}{\omega_2}\tau_{01}\right)$ $\left(\frac{4\omega_3}{\omega_2}\tau_{00}-\alpha, \alpha\right)$ $(1-\alpha, 1-\alpha)$	(α, α) $\left(1-\alpha, 1-\alpha+\frac{4\omega_3}{\omega_2}\tau_{01}\right)$ $(1-\alpha, \alpha)$ $(1-\alpha, 1-\alpha)$
$0 < \tau_{01} \leq \frac{\omega_2}{4\omega_3}(1 - 2\alpha)$	$\left(\alpha, \frac{4\omega_3}{\omega_2}\tau_{01}-\alpha\right)$ $\left(1-\alpha, 1-\alpha+\frac{4\omega_3}{\omega_2}\tau_{01}\right)$ (α, α) $(\alpha, 1-\alpha)$	$\left(\alpha, \frac{4\omega_3}{\omega_2}\tau_{01}-\alpha\right)$ $\left(1-\alpha, 1-\alpha+\frac{4\omega_3}{\omega_2}\tau_{01}\right)$ (α, α) $\left(1-\alpha+\frac{4\omega_3}{\omega_2}\tau_{00}, 1-\alpha\right)$	$\left(\alpha, \frac{4\omega_3}{\omega_2}\tau_{01}-\alpha\right)$ $\left(1-\alpha, 1-\alpha+\frac{4\omega_3}{\omega_2}\tau_{01}\right)$ $\left(\frac{4\omega_3}{\omega_2}\tau_{00}-\alpha, \alpha\right)$ $(1-\alpha, 1-\alpha)$	$\left(\alpha, \frac{4\omega_3}{\omega_2}\tau_{01}-\alpha\right)$ $\left(1-\alpha, 1-\alpha+\frac{4\omega_3}{\omega_2}\tau_{01}\right)$ $(1-\alpha, \alpha)$ $(1-\alpha, 1-\alpha)$
$\frac{\omega_2}{4\omega_3}(1 - 2\alpha) < \tau_{01} \leq 0$	$(\alpha, 1-\alpha)$ $(1-\alpha, 1-\alpha)$ (α, α) $(\alpha, 1-\alpha)$	$(\alpha, 1-\alpha)$ $(1-\alpha, 1-\alpha)$ (α, α) $\left(1-\alpha+\frac{4\omega_3}{\omega_2}\tau_{00}, 1-\alpha\right)$	$(\alpha, 1-\alpha)$ $(1-\alpha, 1-\alpha)$ $\left(\frac{4\omega_3}{\omega_2}\tau_{00}-\alpha, \alpha\right)$ $(1-\alpha, 1-\alpha)$	$(\alpha, 1-\alpha)$ $(1-\alpha, 1-\alpha)$ $(1-\alpha, \alpha)$ $(1-\alpha, 1-\alpha)$

Table 1: List of the 4 candidates at the boundaries to get the optimum

I will refer to the cells as $B(\tau_{00}, \tau_{01})$, a set of 4 points. Whenever $\nexists (x, y) \in [\alpha, 1-\alpha]^2$ such that $\nabla f(x, y) = 0$, one of the points in $B(\tau_{00}, \tau_{01})$ will yield the optimum. In this case, comparing the associated values of $f(x, y)$ is sufficient to find the optimal solution.

Providing each outcome for any (x, g) while treated and not treated, in addition to the preferences for the different dimensions of fairness ω_1, ω_2, ω_3, one can find the optimal allocations using algorithm (1) yielding the highest value for $f(\cdot, \cdot)$. The same procedure can be repeated, changing the treatment effects in the algorithm, to optimize $g(\cdot, \cdot)$. The set of solutions yields a certain utility $U(p^\star)$ that should finally be compared with $U((\pi, \pi, \pi, \pi))$ to determine the best allocation analytically.

Thus, in many cases, there exist several optimal solutions that correspond to intermediate values between $[\alpha, 1-\alpha]$. In some cases, there is an infinite number of solutions. The solutions to the trade-off between the two dimensions are thus not obvious nor unique in general.

[9]Symmetry here arises partly from the uniform distribution of individual characteristics among the population in the simplified framework, although analogous relationship should arise in other contexts. The difference is that with different distributions of the characteristics, the proportionality should be distorted.

3.2.3 Interpreting traditional experimental allocation rules concerning the theoretical model

In certain extreme cases, the choice of experimental design would be imperative to align with the agent's fairness values. Indeed, if the researcher has preferences $\omega_1 = 1$, $\omega_2 = 0$, and $\omega_3 = 0$, then setting up a standard randomized experiment defining π in advance would maximize her fairness objective. If she has a strong preference for increasing impact, i.e. $\omega_1 = 0$, $\omega_2 = 0$, $\omega_3 = 1$, the contextual bandit would be the experimental design to favor. Finally, if she wishes to avoid discriminating according to G but wants to increase the total impact, or if $\omega_1 = 0$, $\omega_2 = 1 - \epsilon$, $\omega_3 = \epsilon$ with $\epsilon \to 0$, then she could opt for a contextual bandit neglecting information on G. I propose to illustrate this idea on simulated data.

4 Simulation in the simplified framework

To illustrate the results above, let us assume some values for the potential outcomes in the simplified framework. For this particular simulation, potential outcomes are specified in figure 1, denoting $\forall (x, g) \in \{0, 1\}^2$ by $Y(1, x, g)$ the observed outcome when an individual of type (x, g) is treated, and $Y(0, x, g)$ when she is not. Thus, in figure 1, each cell represents the potential outcomes of one type of individual, and the corresponding treatment effect can be obtained by substracting in each cell the second potential outcome from the first. The allocation rule that would be pre-determined for transparency is set to $\pi = 0.5$.[10]

4.1 Optimal solution for different preferences

4.1.1 Mixed preferences for the different dimensions of fairness and $\alpha = 0$

First, let the preferences of the agent for the different dimensions of fairness be $\omega_1 = 0.2$, $\omega_2 = 0.7$, $\omega_3 = 0.1$. Also assume in this case that $\alpha = 0$, which means that the need to learn from the experiment is neglected. Since I solved the optimization problem in a more general case, I can already determine the optimal solution of the simplified optimization program by computing the treatment effects in the table. Since $\tau_{00} = 1$ and $\tau_{01} = -2$, we know that there will be no interior solution. The candidates thus can be found on the boundaries $B(-1, 2)$ using table 1. The set of 4 candidates will depend on the agent's preferences for the different dimensions of fairness ω_1, ω_2 and ω_3. Then, we can proceed in the same way for (τ_{10}, τ_{11}) - we can see that the solution will be unique since $\tau_{10} = -1$ and $\tau_{11} = 5$. Moreover, in this setting, $\frac{\omega_2}{4\omega_3} \sim 1.75$. Candidates at the boundary are those in the third column and second row of table 1. Among those candidates, $\left(\frac{4\omega_3}{\omega_2}\tau_{00}, 0\right)$ yields the highest value for $f(\cdot, \cdot)$. Intuitively, the optimal solution is to give sometimes the treatment to the group $(0, 0)$ that has a positive effect, while keeping the distance between p_{00} and p_{01} moderate so that there

[10] Solutions for alternate versions of the model with mixed preferences can be computed using a Shiny app available here: bertillep.shinyapps.io/optim/

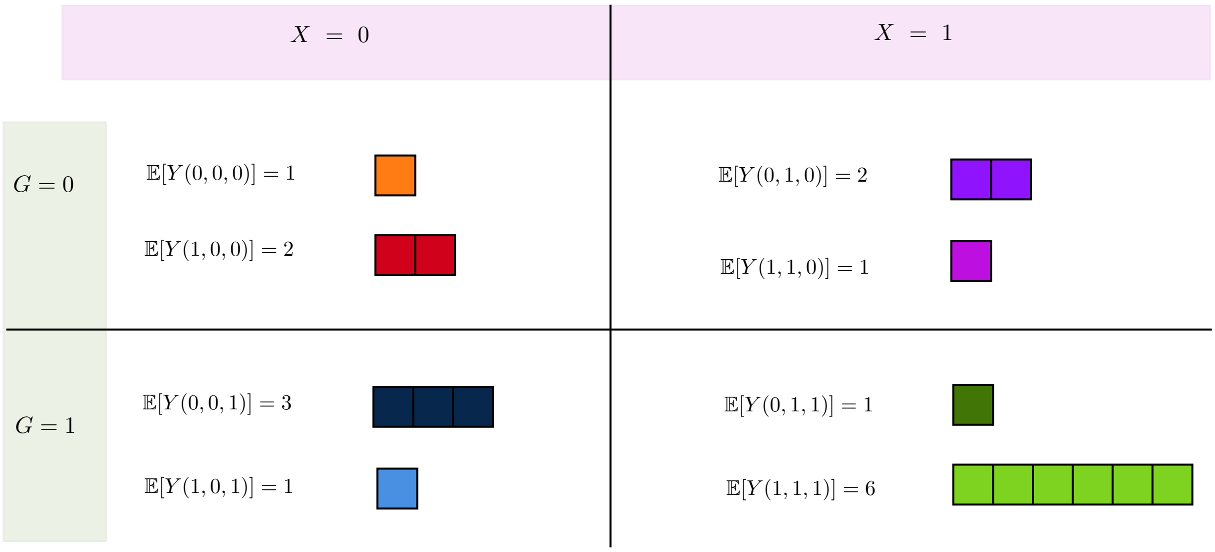

Figure 1: Data Generating Process: - definition of the potential outcomes

Notes: This figure represents the potential outcomes for the four types of individuals when they are treated and when they are not treated. For example, an individual with characteristics $x = 0$ and $g = 1$ would receive on average an outcome of 3 when not treated and 1 when treated. Distribution of the potential outcomes in the sample is plotted in figure 2.

is no huge increase in inequality of opportunity. The distance between the two probabilities is determined by the preferences for the second dimension versus the third dimension of fairness. Similarly, I can deduce that the best choice for (p_{10}, p_{11}) is $\left(\frac{4\omega_3}{\omega_2}\tau_{10}, 1\right)$. Moreover:

$$U\left(\left(\frac{4\omega_3}{\omega_2}\tau_{00}, 0, \frac{4\omega_3}{\omega_2}\tau_{10}, 1\right)\right) > U((\pi, \pi, \pi, \pi))$$

which implies that points maximizing $f(\cdot, \cdot)$ and $g(\cdot, \cdot)$ ignoring the case $p = (\pi, \pi, \pi, \pi)$ will maximize the utility on $[\alpha, 1-\alpha]^4$.

4.1.2 Exclusive preferences for transparency and $\pi \in [\alpha, 1-\alpha]$

Let α be a common trimming threshold (for example, $\alpha = 0.01$ or $\alpha = 0.05$). Let us assume that $\omega_1 = 1$, $\omega_2 = 0$, and $\omega_3 = 0$. Moreover, $\pi \in [\alpha, 1-\alpha]$, meaning the treated and control groups are not totally unbalanced, and the probability of being treated has been chosen so that an important part of the sample is assigned to each group. In this setting, no allocation rule will of course be better than (π, π, π, π) since the other dimensions of fairness do not provide any supplementary satisfaction to the agent. The optimal allocation rule will necessarily align with that of a standard RCT with a predetermined probability of allocation π.

4.1.3 Exclusive preferences for welfare maximization with $\alpha = 0$

Let us assume that $\omega_1 = 0$, $\omega_2 = 0$, and $\omega_3 = 1$. Additionally, assume $\alpha = 0$, meaning that no precautions are taken against extreme sampling probabilities. In this case, the main objective is to maximize the overall impact. The best allocation will be to allocate treatment to individuals with characteristics $(0, 0)$ and $(1, 1)$ only. The objective will align with that of a contextual bandit.

4.1.4 Exclusive preferences for welfare maximization with $\alpha = 0.05$

Let us assume again that $\omega_1 = 0$, $\omega_2 = 0$, and $\omega_3 = 1$. However, let us take some precautions against extreme sampling probabilities by assuming $\alpha = 0.05$. In this case, the main objective is to maximize the overall impact with a constraint: the sampling probabilities must not be less than α or greater than $1 - \alpha$. The best allocation will be to allocate treatment to individuals with characteristics $(0, 0)$ and $(1, 1)$ with probability $1 - \alpha$, and to allocate treatment to the others with probability α. The objective will align with that of a tempered contextual bandit.

4.1.5 Strong preferences for equal opportunity, with marginal interest for welfare maximization

Let us assume that $\omega_1 = 0$, $\omega_2 = 0.99$, and $\omega_3 = 0.01$. This illustrates the case $\omega_1 = 0$, $\omega_2 = 1 - \epsilon$ and $\omega_3 = \epsilon$ with $\epsilon \to 0$. This objective maintains that it is highly costly to give different probabilities to individuals sharing the same characteristic x but with different values for g. However, since the preference for welfare maximization is strictly positive, this means that there is still a concern for welfare maximization. In the extreme case $\epsilon = 0$, any set of probabilities such that $p_{00} = p_{01}$ and $p_{10} = p_{11}$ would maximize the objective. Here, there is a trade-off between keeping the pairs of probabilities very close to each other while still maximizing the overall welfare.

4.2 Choosing an algorithm to converge to different solutions

4.2.1 Description of the 4 experimental designs

Let us illustrate how different algorithms can converge to these different goals in this setting. I generated a sample of 2 000 observations. Each observation has a probability $1/4$ to fall into each category (x, g), which means that each combination is equally represented in the sample. The potential outcomes are defined as in figure 1, adding a small noise from a centered normal distribution with variance 0.5. The histograms of sampled values are displayed in the appendix. The sample is divided into batches of size 100 so that the experiment lasts for 20 periods. Four algorithms are compared:

- A naive "algorithm" or (decision strategy) attributes to each individual at each batch a probability of 0.5 of being treated. This corresponds to a standard RCT design where π would have been set to 0.5.

- A Gaussian Thompson Sampling algorithm with conjugate priors (shortly described in the appendix) uses all the information on the individuals to personalize the treatment. It means that it is allowed to use as well x or g to personalize the treatment. In practice, 4 separate Thompson sampling algorithms are fitted for each of the 4 combinations of characteristics.

- A variant of the previous Gaussian Thompson Sampling algorithm with conjugate priors uses all the information on the individuals to personalize the treatment. The difference is that in this algorithm, a correction is made whenever the sampling probabilities are too extreme (less than α or larger than $1-\alpha$). This algorithm could be interpreted as a tempered version of the previous one and is inspired by the literature on tempered Thompson Sampling (Caria et al. 2020).

- Another Gaussian Thompson Sampling algorithm with conjugate priors uses only the information on x and is not allowed to discriminate using g. That means that for a given x, the same decision should be taken for both $g = 0$ and $g = 1$. In practice, two separate Thompson sampling algorithms are fitted for $x = 0$ and $x = 1$.

4.2.2 Simulations results

The respective behaviors of these three algorithms are illustrated in figures 3, 4, 5 and 6. This section aims to interpret these simulations results.

An RCT reaches the optimal solution for exclusive preferences for transparency and $\pi \in [\alpha, 1-\alpha]$. Obviously, the probabilities set up in advance to 0.5 maximize the program with $\omega_1 = 1$, $\omega_2 = 0$ and $\omega_3 = 0$ (figure 3). The solution is displayed on the graph as a purple horizontal line corresponding to the value 0.5. This shows that an RCT may be aligned with the solution of the fairness objective with an extreme parametrization, where only transparency matters.

A Gaussian Thompson Sampling taking into account information on both variables reaches the optimal solution for exclusive preferences for welfare maximization with $\alpha = 0$ (figure 4). Using the second decision strategy, that is optimizing using all the information on the individuals, the probabilities converge to the optimal probabilities when $\omega_1 = 0$, $\omega_2 = 0$ and $\omega_3 = 1$, that is, when there is an exclusive preference for welfare maximization. The solution is displayed on this graph as a red line and takes values 0 or 1 depending on whether being assigned to the treatment group yields the highest outcome or not.

A Tempered Gaussian Thompson Sampling taking into account information on both variables reaches the optimal solution for exclusive preferences for welfare maximization with $\alpha = 0.1$ (figure 4). The solutions are displayed as a green line on each graph.

Finally, using only information on x to optimize the allocation is a good way to get closer to the optimal solution when $\omega_1 = 0$, $\omega_2 = 1 - \epsilon$ and $\omega_3 = \epsilon$ with $\epsilon \to 0$ (figure 6).

Thus, these four experimental designs align with very extreme interpretations of the fairness objective. None of these designs can reach the fairness objective with mixed preferences ($\omega_1 = 0.2$, $\omega_2 = 0.7$, $\omega_3 = 0.1$).

5 Conclusion

This paper presents a simple model that still addresses various fairness concerns in experimental design by drawing from the economic literature on fairness. It emphasizes the crucial role of treatment effects in determining optimal allocation strategies. For instance, when there is even a slight consideration for maximizing welfare and all treatment effects are positive, treating all individuals becomes the optimal choice. The paper also highlights scenarios in which treatment benefits some individuals while being detrimental to others. In such cases, a multitude of optimal solutions may exist, underscoring the intricate trade-offs involved in balancing different fairness dimensions.

Moreover, a simulation study demonstrates how basic experimental designs can converge toward optimal solutions that embody radical interpretations of fairness. In other words, preferences for RCT or adaptive experiments could arise from different radical interpretations of fairness.

These findings raise important questions about the acceptability and fairness of experimental designs. Firstly, it prompts a discussion on how to establish fairness preferences to design experiments that are acceptable to both institutions and experimental subjects. Secondly, it calls for the exploration of alternative experimental designs that can accommodate mixed preferences and yield interpretations aligned with the fairness objectives.

References

Angwin, J., J. Larson, S. Mattu, and L. Kirchner (2016): "Machine Bias," Retrieved on June 24, 2023. Available at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Athey, S., K. Bergstrom, V. Hadad, J. C. Jamison, B. Ozler, L. Parisotto, J. D. Sama, et al. (2021): "Shared Decision-Making: Can Improved Counseling Increase Willingness to Pay for Modern Contraceptives?" .

Banerjee, A. V., E. Duflo, and M. Kremer (2016): "The influence of randomized controlled trials on development economics research and on development policy," *The state of Economics, the state of the world*, 439–498.

Bigman, Y. E. and K. Gray (2018): "People are averse to machines making moral decisions," *Cognition*, 181, 21–34.

Blinder, A. S. (1973): "Wage discrimination: reduced form and structural estimates," *Journal of Human resources*, 436–455.

Bubeck, S. and N. Cesa-Bianchi (2012): "Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems," *Foundations and Trends® in Machine Learning*, 5, 1–122.

Caria, S., M. Kasy, S. Quinn, S. Shami, A. Teytelboym, and others (2020): "An adaptive targeted field experiment: Job search assistance for refugees in Jordan," *Publisher: CESifo Working Paper.*

Chapelle, O. and L. Li (2011): "An empirical evaluation of thompson sampling," *Advances in neural information processing systems*, 24.

Chugunova, M. and W. J. Luhan (2022): "Ruled by robots: Preference for algorithmic decision makers and perceptions of their choices," *Max Planck Institute for Innovation & Competition Research Paper.*

Cowgill, B. and C. E. Tucker (2019): "Economics, fairness and algorithmic bias," *preparation for: Journal of Economic Perspectives.*

Dastin, J. (2018): "Amazon scraps secret AI recruiting tool that showed bias against women," Retrieved on June 24, 2023. Available at https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

Dur, R., A. Non, P. Prottung, and B. Ricci (2023): "Who's Afraid of Policy Experiments?" .

FISHER, R. A. (1925): *Statistical Methods for Research Workers*, Genesis Publishing Pvt Ltd.

FLACHAIRE, E. AND B. PICARD (2023): "Inequality decomposition with machine learning," *Unpublished*.

GOGOLL, J. AND M. UHL (2018): "Rage against the machine: Automation in the moral domain," *Journal of Behavioral and Experimental Economics*, 74, 97–103.

IMBENS, G. W. AND D. B. RUBIN (2015): *Causal Inference for Statistics, Social, and Biomedical Sciences*, Cambridge University Press.

KITAGAWA, E. M. (1955): "Components of a difference between two rates," *Journal of the american statistical association*, 50, 1168–1194.

KNAUS, M. C., M. LECHNER, AND A. STRITTMATTER (2021): "Machine learning estimation of heterogeneous causal effects: Empirical Monte Carlo evidence," *The Econometrics Journal*, 24, 134–161.

MIRACLE, V. A. (2016): "The Belmont Report: The triple crown of research ethics," *Dimensions of critical care nursing*, 35, 223–228.

NEUMARK, D. (1988): "Employers' discriminatory behavior and the estimation of wage discrimination," *Journal of Human resources*, 279–295.

NEYMAN, J. (1990): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," *Statistical Science*, 5, 465–472, translated by Dorota M. Dabrowska and Terence P. Speed.

OAXACA, R. (1973): "Male-female wage differentials in urban labor markets," *International economic review*, 693–709.

OFFER-WESTORT, M., A. COPPOCK, AND D. P. GREEN (2021): "Adaptive experimental design: Prospects and applications in political science," *American Journal of Political Science*, 65, 826–844.

RUBIN, D. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.

VIVIANO, D. AND J. BRADIC (2022): "Fair Policy Targeting," .

WACHTER, S., B. MITTELSTADT, AND C. RUSSELL (2021): "Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI," *Computer Law & Security Review*, 41, 105567.

ZHAN, R., V. HADAD, D. A. HIRSHBERG, AND S. ATHEY (2021): "Off-Policy Evaluation via Adaptive Weighting with Data from Contextual Bandits," .

A Proofs

Simplified model

$$
\begin{aligned}
U(p) = \quad & \omega_1 \mathbb{1}\{p_{00}=\pi\} \mathbb{1}\{p_{01}=\pi\} \mathbb{1}\{p_{10}=\pi\} \mathbb{1}\{p_{11}=\pi\} \\
& -\omega_2 \left(\left(p_{00} - \frac{p_{00}+p_{01}}{2} \right)^2 \left(-\frac{1}{2} \right)^2 \right. \\
& + \left(p_{01} - \frac{p_{00}+p_{01}}{2} \right)^2 \left(\frac{1}{2} \right)^2 \\
& + \left(p_{10} - \frac{p_{10}+p_{11}}{2} \right)^2 \left(-\frac{1}{2} \right)^2 \\
& + \left(p_{11} - \frac{p_{10}+p_{11}}{2} \right)^2 \left(-\frac{1}{2} \right)^2 \Bigg) \\
& \left. +\omega_3 \sum_{(k,1)\in\{0,1\}^2} \left(p_{kl}\tau_{kl} + c_{kl} \right) \right) \\
= \quad & \omega_1 \mathbb{1}\{p_{00}=\pi\} \mathbb{1}\{p_{01}=\pi\} \mathbb{1}\{p_{10}=\pi\} \mathbb{1}\{p_{11}=\pi\} \\
& -\frac{\omega_2}{4} \left(2 \times \left(\frac{p_{00}-p_{01}}{2} \right)^2 + 2 \times \left(\frac{p_{10}-p_{11}}{2} \right)^2 \right) \\
& + \omega_3 \sum_{(k,1)\in\{0,1\}^2} \left(p_{k\ell}\tau_{kl} + c_{kl} \right) \\
= \quad & \omega_1 \mathbb{1}\{p_{00}=\pi\} \mathbb{1}\{p_{01}=\pi\} \mathbb{1}\{p_{10}=\pi\} \mathbb{1}\{p_{11}=\pi\} \\
& -\frac{\omega_2}{8} \left(\left(p_{00}-p_{01} \right)^2 + \left(p_{10}-p_{11} \right)^2 \right) \\
& + \omega_3 \sum_{(k,l)\in\{0,1\}^2} \left(p_{kl}\tau_{kl} + c_{kl} \right)
\end{aligned}
$$

Optimal solutions in the simplified model for mixed preferences

Let us consider first every possible solutions $p^* \neq (\pi,\pi,\pi,\pi)$, assuming $\omega_2 \neq 0$ and $\omega_3 \neq 0$, for $p \neq (\pi,\pi,\pi,\pi)$:

$$
\begin{aligned}
U(p) = & \frac{-\omega_2}{8} \left(\left(p_{00}-p_{01} \right)^2 + \left(p_{10}-p_{11} \right)^2 \right) \\
& + \omega_3 \sum_{(k,l)\in\{0,1\}^2} \left(\tau_{kl} p_{kl} + c_{kl} \right)
\end{aligned}
$$

Then the function can be decomposed into two separate functions to optimize and a constant:

$$
U(p) = f\left(p_{00}, p_{01}\right) + g\left(p_{10}, p_{11}\right) + \text{ constant}
$$

Where:

$$
\begin{aligned}
f(x,y) &= -\frac{\omega_2}{8}(x-y)^2 + \omega_3 \left(x\tau_{00} + y\tau_{01} \right) \\
g(x,y) &= -\frac{\omega_2}{8}(x-y)^2 + \omega_3 \left(x\tau_{10} + y\tau_{11} \right)
\end{aligned}
$$

The two independent maximization problems are very similar. Let us focus on the maximization of $f(x,y)$ for $(x,y) \in [\alpha, 1-\alpha]^2$

Boundary solutions

Let us compute the points maximizing the function at the boundaries, fixing $x = \alpha$, $x = 1-\alpha$, $y = \alpha$, and finally $y = 1-\alpha$.

- **Case 1:** $x = \alpha$

$$\begin{aligned}
&f(\alpha, y) = -\frac{\omega_2}{8}(\alpha - y)^2 + \omega_3 y \tau_{01} \\
&\frac{\partial}{\partial y} f(\alpha, y) = -\frac{\omega_2}{4}(\alpha - y) + \omega_3 \tau_{01} \geq 0 \\
&\Leftrightarrow \frac{\omega_2}{4} y - \frac{\omega_2}{4} \alpha \leq \omega_3 \tau_{01} \\
&\Leftrightarrow y \leq \alpha + \frac{4\omega_3}{\omega_2} \tau_{01}
\end{aligned}$$

case 1a: $\tau_{01} < 0 \Rightarrow \forall y, \frac{\partial}{\partial y} f(\alpha, y) \leqslant 0$

$$\Rightarrow y^* = \alpha$$

case 1b $\begin{cases} \alpha + \frac{4\omega_3}{\omega_2}\tau_{01} \geq \alpha \Leftrightarrow \tau_{01} \leq 0 \\ \alpha + \frac{4\omega_3}{\omega_2}\tau_{01} \leq 1 - \alpha \Leftrightarrow \tau_{01} \leq \frac{\omega_2}{4\omega_3} \end{cases}$
Then $y^* = \alpha + \frac{4\omega_3}{\omega_2}\tau_{01}$

case 1c $\alpha + \frac{4\omega_3 \tau_{01}}{\omega_2} > 1 - \alpha \quad \Leftrightarrow \tau_{01} > \frac{\omega_2}{4\omega_3}(1 - 2\alpha)$
Then $y^* = 1 - \alpha$

$$B_1 = \begin{cases} (\alpha, \alpha) & \text{if } \tau_{01} < 0 \\ \left(\alpha, \frac{4\omega_3}{\omega_2}\tau_{01} - \alpha\right) & \text{if } \tau_{01} \in \left[0, \frac{\omega_2}{4\omega_3}(1 - 2\alpha)\right] \\ (\alpha, 1 - \alpha) & \text{if } \tau_{01} > \frac{\omega_2}{4\omega_3}(1 - 2\alpha) \end{cases}$$

- **Case 2,** $x = 1 - \alpha$

$$
\begin{aligned}
&f(1-\alpha, y) = -\frac{\omega_2}{8}(1-\alpha-y)^2 + \omega_3\left(\tau_{00} + y\tau_{01}\right) \\
&\frac{\partial f}{\partial y}(1-\alpha, y) = \frac{\omega_2}{4}(1-\alpha-y) + \omega_3\tau_{01} \geq 0 \\
&\Leftrightarrow \quad 1-\alpha-y \geq -\frac{4\omega_3}{\omega_2}\tau_{01} \\
&\Leftrightarrow \quad -y \geq -\frac{4\omega_3}{\omega_2}\tau_{0_1} - 1 + \alpha \\
&\Leftrightarrow \quad y \leq \frac{4\omega_3}{\omega_2}\tau_{01} + 1 - \alpha
\end{aligned}
$$

case 2a $\quad \tau_{01} > 0$ then $\forall y \quad \frac{\partial}{\partial y} f(1-\alpha, y) \geq 0$
thus $y^* = 1-\alpha$

case 2b

$$
\left\{\begin{array}{l}
1-\alpha+4\frac{\omega_3}{\omega_2}\tau_{01} \geq \alpha \\
1-\alpha+\frac{4\omega_3}{\omega_2}\tau_{01} \leq 1-\alpha
\end{array}\right. \Leftrightarrow \left\{\begin{array}{l}
\tau_{01} \geq \frac{\omega_2}{4\omega_3}(2\alpha-1) \\
\tau_{01} \leq 0
\end{array}\right.
$$

Then $y^* = \frac{4\omega_3}{\omega_2}\tau_{01} + 1$

case 2c $\quad \tau_{01} < \frac{\omega_2}{4\omega_3}(2\alpha-1)$
Then $y^* = \alpha$

$$
B_2 = \begin{cases}
(1-\alpha, \alpha) & \text{if } \tau_{01} < \frac{\omega_2}{4\omega_3}(2\alpha-1) \\
\left(1-\alpha, 1-\alpha+\frac{4\omega_3}{\omega_2}\tau_{01}\right) & \text{if } \frac{\omega_2}{4\omega_3}(2\alpha-1) \leq \tau_{01} \leq 0 \\
(1-\alpha, 1-\alpha) & \text{if } \tau_{01} \geq 0
\end{cases}
$$

- **Case 3:** $y = \alpha$

This case is symmetrical to case 1:

$$
B_3 = \begin{cases}
(\alpha, \alpha) & \text{if } \tau_{00} < 0 \\
\left(\frac{4\omega_3}{\omega_2}\tau_{00} - \alpha, \alpha\right) & \text{if } 0 \leq \tau_{00} \leq \frac{\omega_2}{4\omega_3}(1-2\alpha) \\
(1-\alpha, \alpha) & \text{if } \tau_{00} > \frac{\omega_2}{4\omega_3}(1-2\alpha)
\end{cases}
$$

- **Case 4:** $y = 1-\alpha$

This case is symmetrical to case 2:

$$B_4 = \begin{cases} (\alpha, 1-\alpha) & \text{if } \tau_{00} \leq \frac{\omega_2}{4\omega_3}(2\alpha - 1) \\ 1-\alpha + \left(\frac{4\omega_3}{\omega_2}\tau_{00}, 1-\alpha\right) & \text{if } \frac{\omega_2}{4\omega_3}(2\alpha-1) \leq \tau_{00} \leq 0 \\ (1-\alpha, 1-\alpha) & \text{if } \tau_{00} \geq 0 \end{cases}$$

Interior solutions

Given $(x, y) \in [\alpha, 1-\alpha]^2, \quad \nabla f(x, y) = 0$ is equivalent to:

$$\begin{cases} -\frac{\omega_2}{4}(x-y) + \omega_3\tau_{00} = 0 \quad (1) \\ \frac{\omega_2}{4}(x-y) + \omega_3\tau_{01} = 0 \quad (2) \end{cases}$$

$(1) + (2) \Rightarrow \omega_3(\tau_{00} + \tau_{01}) = 0$

- Thus, if $\omega_3 \neq 0$, the above system is equivalent to

$$\begin{cases} -\frac{\omega_2}{4}(x-y) + \omega_3\tau_{00} = 0 \quad (1') \\ \tau_{00} = -\tau_{01} \quad (2') \end{cases}$$

There is only one possible configuration if $\omega_3 \neq 0$ for having interior solutions: the treatment effects should be symmetrical.

Then, if $\tau_{00} \geq 0$

$(1') \Leftrightarrow x = \frac{4\omega_3}{\omega_2}\tau_{00} + y$

$(1')$ and $(x \in [\alpha, 1-\alpha])$ and $(y \in (\alpha, 1-\alpha])$

$\Leftrightarrow (1')$ and $\left(\frac{4\omega_3}{\omega_2}\tau_{00} \leq 1-\alpha-y\right)$ and $(y \in [\alpha, 1-\alpha])$

$\Leftrightarrow (1')$ and $\left(y \leq 1-\alpha-\frac{4\omega_3}{\omega_2}\tau_{00}\right)$ and $(y \in [\alpha, 1-\alpha])$

$\Leftrightarrow (1')$ and $\left(y \in \left[\alpha, \min\left(1-\alpha, 1-\alpha-\frac{4\omega_3}{\omega_2}\tau_{00}\right)\right]\right)$

Provided that $\frac{4\omega_3}{\omega_2}\tau_{00} \leq 1-2\alpha$

$$S = \left\{x = \frac{4\omega_3}{\omega_2}\tau_{00} + y, y \in \left[\alpha, 1-\alpha-\frac{4\omega_3}{\omega_2}\tau_{00}\right]\right\}$$

If $\frac{4\omega_3}{\omega_2}\tau_{00} > 1$, there are no interior solution

If $\tau_{00} \leq 0$

(1) and $(x \in [\alpha, 1-\alpha])$ and $(y \in [\alpha, 1-\alpha])$

Then:

$$S = \left\{x = \frac{4\omega_3}{\omega_2}\tau_{00} + y, y \in \left[\alpha - \frac{4\omega_3}{\omega_2}\tau_{00}, 1-\alpha\right]\right\}$$

Provided that $\frac{-4\omega_3}{\omega_2}\tau_{00} < 1-2\alpha$

- If $\omega_3 = 0$, $S = \left\{(x, y) \in [\alpha, 1-\alpha]^2, x = y\right\}$

Algorithm 1 Find Optimum for $f(\cdot,\cdot)$

```
procedure FindOptimum(τ00, τ01, ω2, ω3)
    if τ00 = −τ01 and (((4 · ω3 · τ00/ω2 ≤ 1 − 2α) ∧ (τ00 ≥ 0)) ∨ (τ00 < 0 ∧ −4 · ω3 · τ00/ω2 ≤ 1 − 2α)) then
        Print "There exist (x, y) ∈ [α, 1 − α]^2 such that the gradient is zero"
        if ω3 = 0 then
            Print "Any (x, y) ∈ [α, 1 − α]^2 such that x = y provides the maximum"
        else
            if τ00 ≥ 0 then
                m ← 4 · ω3 · τ00/ω2
                Print "Any (x, y) such that x = m + y, y ∈]α, 1 − α − m[ provides the maximum"
            else
                m ← 4 · ω3 · τ00/ω2
                Print "Any (x, y) such that x = m + y, y ∈]α − m, 1 − α[ provides the maximum"
            end if
        end if
    else
        Print "No interior solution, checking the boundaries"
        boundary_points ← B(τ00, τ01, ω2, ω3, α)
        max_sub_funct ← −∞
        for each couple in boundary_points do
            sub_funct_value ← −(ω2/8)(couple[0] − couple[1])^2 + ω3(τ00 x + τ01 y)
            if sub_funct_value > max_sub_funct then
                max_sub_funct ← sub_funct_value
                best_couple ← couple
            end if
        end for
        Print "Point yielding the maximum is {best_couple}"
    end if
end procedure
```

B Gaussian Thompson sampling with conjugate priors

Assuming $\omega_k = (m_k, v_k, \alpha_k, \beta_k)$ with $\alpha_k > 0$ and $\beta_k > 0$

$y_i(k) | \, \mu_k, \sigma_k^2 \sim \mathcal{N}\left(\mu_k, \sigma_k^2\right)$ for all $i \mid t_i = k$

$\left(\mu_k, \sigma_k^2\right) \sim \text{N-}\Gamma^{-1}\left(\omega_k\right)$

$$\mathbb{P}\left(\mu_k, \sigma_k^2 \mid Y(k)\right) = \frac{\mathbb{P}\left(Y(k) \mid \mu_k, \sigma_k^2\right)\mathbb{P}\left(\mu_k, \sigma_k^2\right)}{\mathbb{P}\left(Y_k\right)}$$

$$\propto \mathbb{P}\left(Y(k) \mid \mu_k, \sigma_k^2\right)\mathbb{P}\left(\mu_k, \sigma_k^2\right)$$

The posterior $f_{\left(\mu_k, \sigma_k^2\right)|y(k)}\left(m, s^2 \mid y(k)\right)$ is thus is proportional to :

$$\left(\prod_{i=1,t_i=k}^{t-1} f_{Y_i(k)|\mu_k,\sigma_k^2}\left(y_i(k) | \, m, s^2\right)\right)\text{pdf}_{\text{N-}\Gamma^{-1}}\left(m, s^2\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma_k}\right)^{N_k(t-1)} \exp\left(\frac{\sum_{i=1,t_i=k}^{t-1} - \left(y_i(k) - \mu_k\right)^2}{2\sigma_k^2}\right)\text{pdf}_{\text{N-}\Gamma^{-1}}\left(m, s^2\right)$$

where $N_k(t-1) = \sum_{i=1,t_i=k}^{t-1} \mathbb{1}\left(t_i = k\right), \quad y(k) = \left(y_i(k)\right)_{i\in[1,t-1]}$

$$\text{pdf}_{\text{N-}\Gamma^{-1}}(m, s) = \frac{\sqrt{\nu_k}}{\sqrt{2\pi s^2}}\frac{\beta_k^{\alpha_k}}{\Gamma(\alpha_k)}\left(\frac{1}{s^2}\right)^{\alpha_{k+1}} \exp\left(-\frac{2\beta_k + \nu_k(m - m_k)^2}{2s^2}\right)$$

From this, I get[11]:

$$\mathbb{P}\left(\mu_k, \sigma_k^2 \mid Y(k)\right) \propto \mathcal{N}\left(\mu_k; \sigma_k^2, \omega_k\right)\mathcal{IG}\left(\sigma_k^2; \omega_k\right)$$

[11] See https://gertjanvandenburg.com/blog/thompson_sampling/ for detailed calculations

Algorithm 2 Gaussian Thompson Sampling with Conjugate Priors

Input: α, β, ν, m - all arrays of size equal to the number of arms
Output: Choosen arm and update the parameters
procedure CHOOSE_ARM
 $mu \leftarrow []$
 for i in range(num_arms) **do**
  Draw $sigma\_square$ from inverse gamma distribution
  $zeta\_square[i] \leftarrow \frac{sigma\_square}{Nt[i]+\nu[i]}$
  Append $zeta\_square[i]$ to $history\_zeta\_square[i]$
  $mu\_a \leftarrow$ draw from normal distribution with mean $rot[i]$ and standard deviation $\sqrt{zeta\_square[i]}$
  Append mu_a to mu
  Append $rot[i]$ to $history\_rot[i]$
 end for[12]
 $a \leftarrow$ index of maximum element in mu
 Append a to all_chosen_arm
 return a
end procedure
procedure UPDATE_PARAMETERS(a, xt)
 $old\_xbar\_t \leftarrow xbar\_t[a]$
 Append xt to $all\_xt[a]$
 $xbar\_t[a] \leftarrow old\_xbar\_t + \frac{1}{Nt[a]+1} \cdot (xt - old\_xbar\_t)$
 $st[a] \leftarrow st[a] + xt^2 + Nt[a] \cdot old\_xbar\_t^2 - (Nt[a]+1) \cdot xbar\_t[a]^2$
 $Nt[a] \leftarrow Nt[a] + 1$
 $rot[a] \leftarrow \frac{\nu[a] \cdot m[a] + Nt[a] \cdot xbar\_t[a]}{\nu[a]+Nt[a]}$
 $beta[a] \leftarrow init\_beta[a] + \frac{st[a]}{2} + \frac{Nt[a] \cdot \nu[a] \cdot (xbar\_t[a]-m[a])^2}{2 \cdot (Nt[a]+\nu[a])}$
 $alpha[a] \leftarrow init\_alpha[a] + \frac{Nt[a]}{2}$
 Append $Nt[a]$ to $all\_Nt[a]$
end procedure

[11] In the tempered version of the algorithm, one step is added. If rot and $zeta\_square$ were such that one of the treatment arms would yield the highest mu_a with a probability larger than $1-\alpha$, then rewards mu are drawn again. This is done using the fact that all the mu_a are drawn from two independent normal distributions. Thus, the probability that one is yielding the highest mu is easy to compute using the parameters of the two distributions. To ensure that the probability of drawing this arm remains equal to $1-\alpha$, I draw mu from two normal distributions such that one is yielding the highest mu with exact probability $1-\alpha$.

C Figures

Figure 2: Sampling values for the potential outcomes

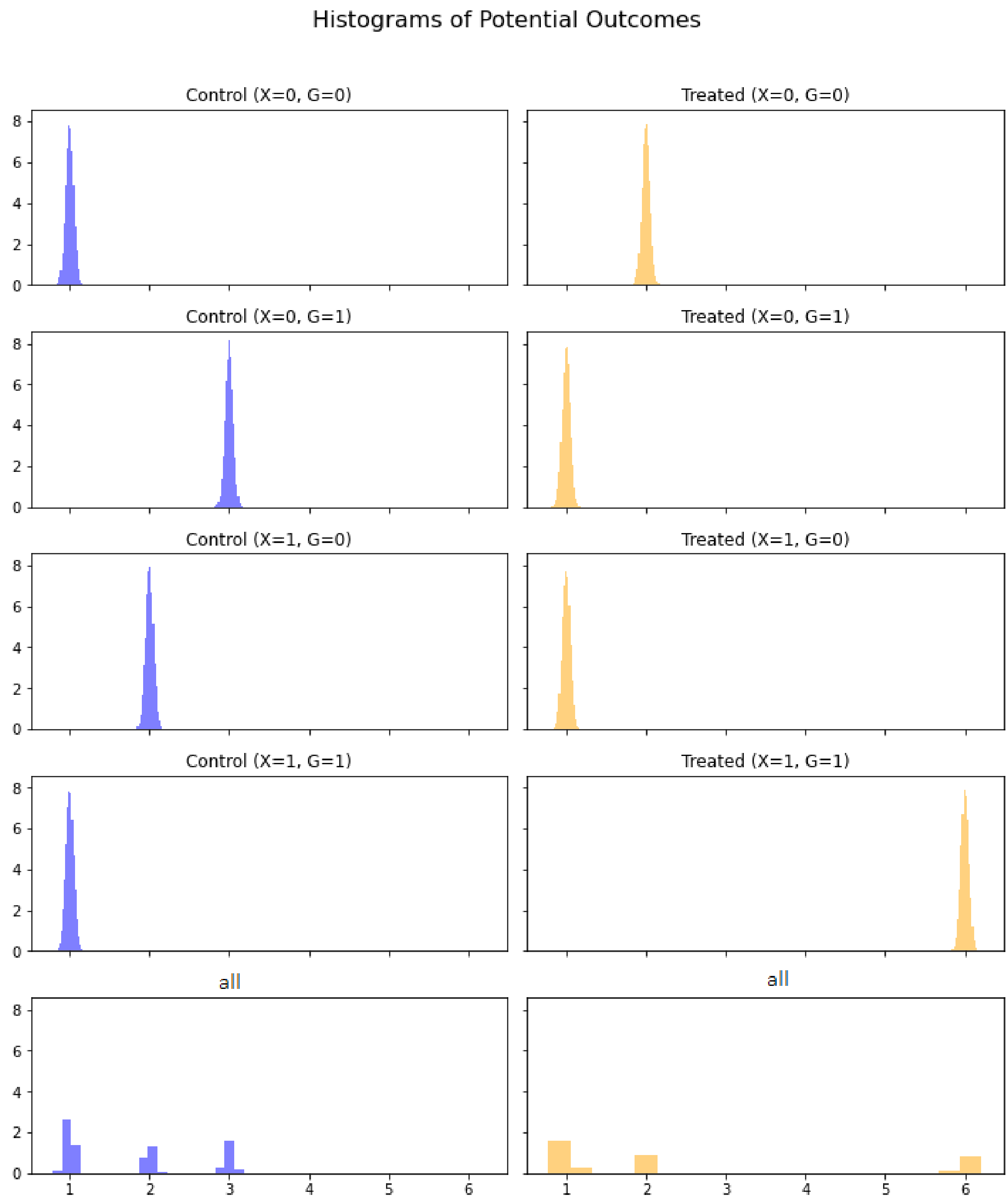

Figure 3: Probabilities in a standard RCT

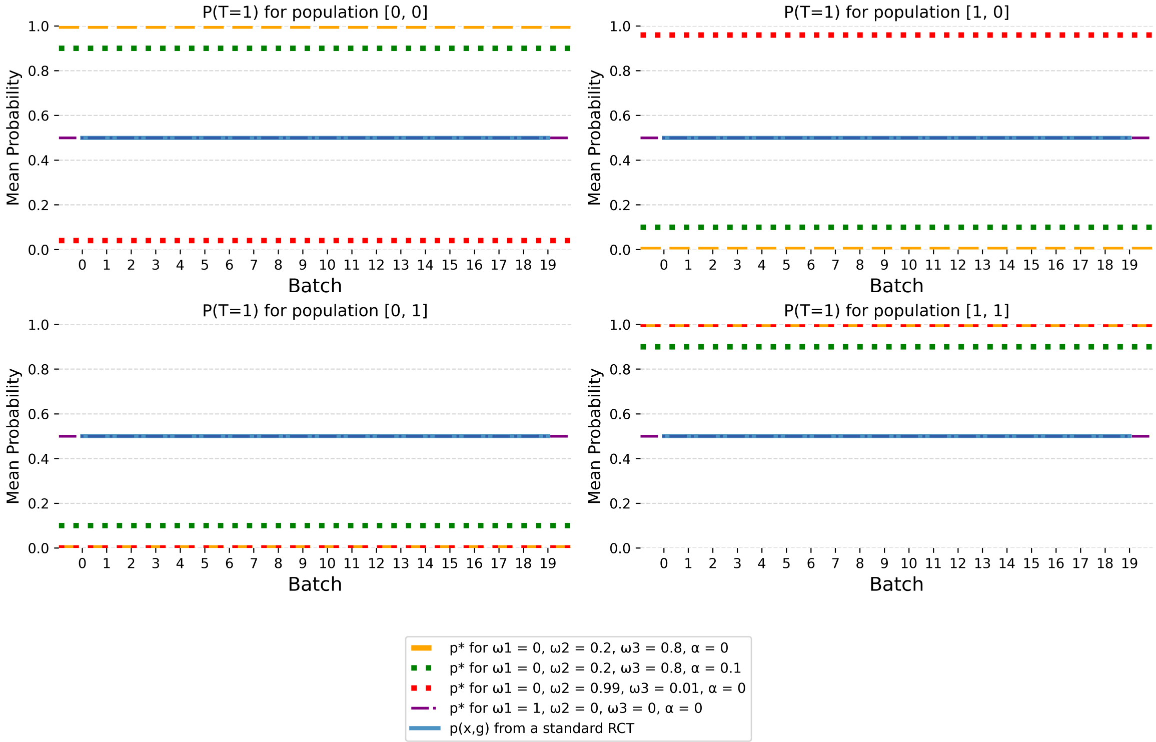

Figure 4: Probabilities using information on x and g

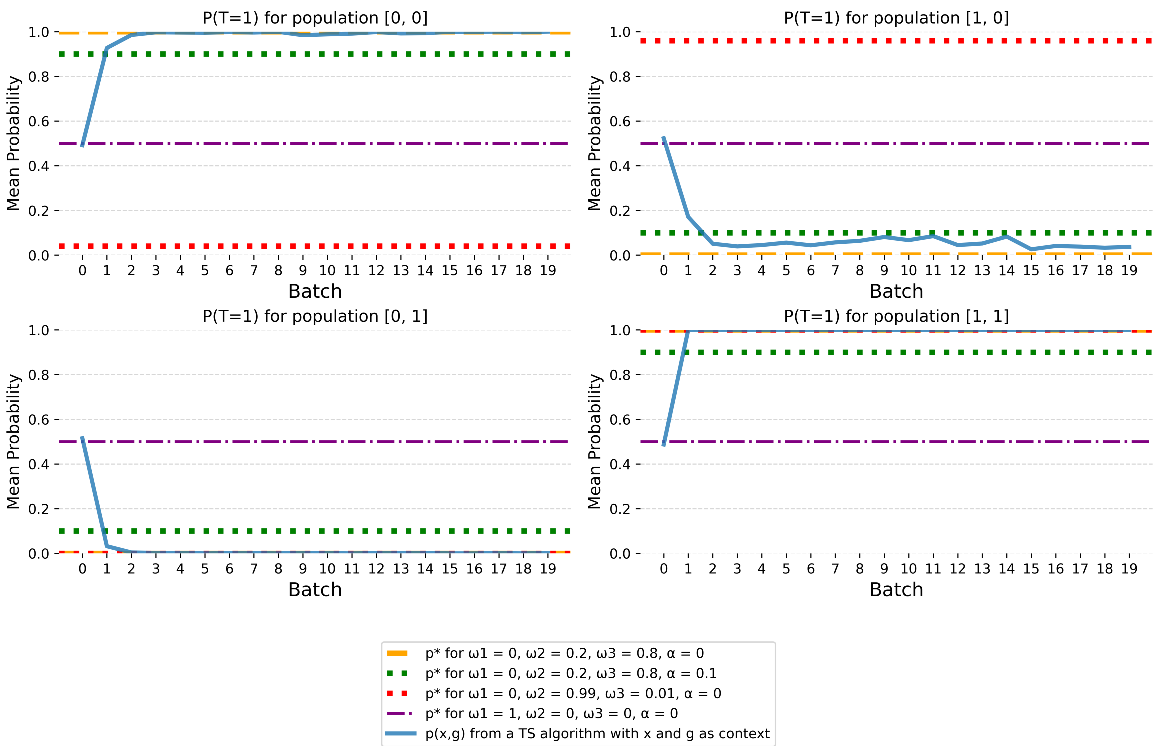

Figure 5: Probabilities using information on x and g

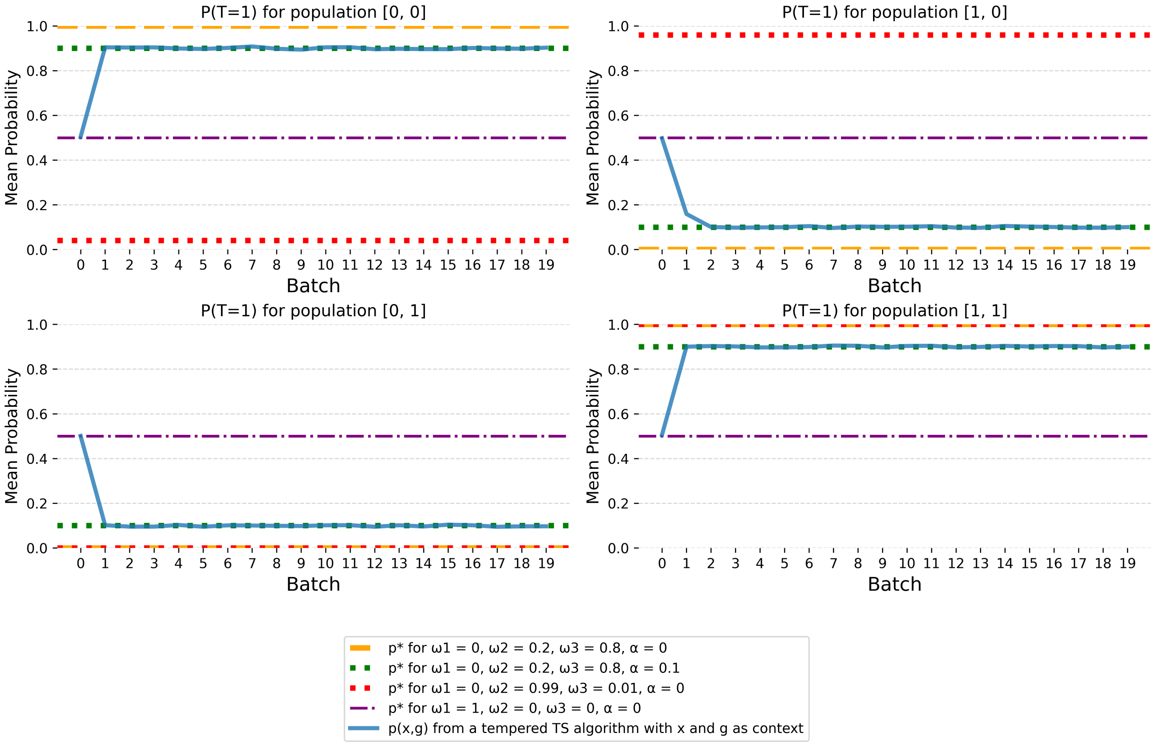

Figure 6: Probabilities using only information on x