The Power of Incentives

Aleksandr Alekseev*
University of Regensburg

July 14, 2022

Abstract

I study an optimal design of monetary incentives in experiments where incentives are a treatment variable. I introduce the Budget Minimization problem in which a researcher chooses the level of incentives that allows her to detect a predicted treatment effect while minimizing her expected budget. The Budget Minimization problem builds upon the power analysis and structural modeling. It extends the standard optimal design approach by explicitly making the budget a part of the objective function. I show theoretically that the problem has an interior solution under fairly mild conditions. I illustrate the applications of the Budget Minimization problem using existing experiments and offer a practical guide for implementing it. My approach adds to the experimental economists' toolkit for an optimal design, however, it also challenges some conventional design recommendations.

Keywords: incentives, economic experiments, experimental design, power analysis, sample size, effect size

JEL Codes: C9, D9

*Department of Economics, University of Regensburg, Universitätsstr. 31, 93040 Regensburg, Germany, e-mail: aleksandr.alekseev@ur.de, phone: + 49 941 943 2740, ORCID: 0000-0001-6542-1920.

1 Introduction

Incentives are a cornerstone of experimental economics. The three main methodological questions about the use of incentives are: whether subjects should be paid, how much subjects should be paid, and how subjects should be paid. Over the years, the field has accumulated a voluminous empirical literature in an attempt to inform the answers to these questions. The findings of that literature are mixed.[1] The theoretical work, on the other hand, has been scarce. Following the early contributions on the first question on whether to pay subjects (Smith, 1976, 1982), the recent literature has mostly been occupied with the third question about how to pay subjects, or incentive compatibility of payoff mechanisms (Cox, Sadiraj, and Schmidt, 2014; Azrieli, Chambers, and Healy, 2018, 2020; Li, 2021). The second question about how much to pay subjects so far has received no theoretical treatment. I attempt to fill in this gap by offering three main contributions. First, I use a simple utility-based framework to formalize the question about the optimal level of incentives in case when incentives are a treatment variable. Second, I show theoretically that this question is well-defined under fairly mild conditions. Third, I illustrate my approach using the data from several existing experiments and offer a practical guide for implementing it.

The current approach to how much to pay subjects is typically ad hoc. It usually amounts to setting incentives at some conventional level based on past experiments, a target hourly wage, or lab policies. None of these conventions, however, are standard within the field (Cloos, Greiff, and Rusch, 2020). In the trivial case when subjects' behavior is not sensitive to incentives, it appears reasonable to set them at the lowest possible value.[2] In many cases, however, subjects' behavior *is* sensitive to incentives, and the question about the optimal level becomes non-trivial.[3] The focus

[1] For reviews, see Camerer and Hogarth (1999), Hertwig and Ortmann (2001), Gneezy, Meier, and Rey-Biel (2011), Cox and Sadiraj (2019), and Voslinsky and Azar (2021). See Kahneman and Tversky (1979) for an early example of the argument in favor of hypothetical incentives. Some of the examples of studies showing no difference in behavior under monetary and hypothetical incentives include Grether and Plott (1979) (preference reversals), Amir, Rand, and Gal (2012) (ultimatum, trust, and public goods games), Brañas-Garza, Kujal, and Lenkei (2019) (cognitive reflection test), and Enke et al. (2021) (cognitive reflection test, contingent reasoning, base rate neglect, and anchoring). The counter-examples include Smith and Walker (1993) (auctions), Cox and Grether (1996) (preference reversals), Holt and Laury (2002) (risk preferences), Amir, Rand, and Gal (2012) (dictator game), and Rousu et al. (2015) (classroom experiments on prisoner's dilemma), Kleinlercher and Stöckl (2018) (finance experiments).

[2] Indeed, some papers show no difference in behavior between small and large incentives: Fischbacher and Föllmi-Heusi (2013) (dishonest behavior), Parravano and Poulsen (2015) (asymmetric coordination games), Thielmann, Heck, and Hilbig (2016) (trust game), Araujo et al. (2016) (real-effort task), Pulford, Colman, and Loomes (2018) (various two-player 3x3 and 4x4 games), and Larney, Rotella, and Barclay (2019) (meta-analysis of incentives in ultimatum and dictator games).

[3] The examples of papers showing a difference in behavior between small and large incentives include Gibson, Tanner, and Wagner (2013); Balasubramanian, Bennett, and Pierce (2017) (dishonest behavior), El Harbi et al.

of the present paper is on the latter cases, in particular, when incentives are a treatment variable.

A key factor that enables studying the optimal level of incentives is that researchers are often interested in testing qualitative hypotheses. A typical research question is whether a treatment variable affects subjects' behavior while the specific values of the treatment variable are nuisance parameters.[4] For example, a researcher studying performance pay is more likely to be interested in whether a higher piece rate increases effort rather than whether a specific 2-cent bump in a piece rate increases effort.[5] The qualitative nature of hypotheses creates a degree of freedom that I exploit to pick an "optimal," in a sense precisely defined below, level of incentives.

I introduce a *Budget Minimization problem* in which a researcher chooses the level of monetary incentives that allows her to find a predicted treatment effect for some conventional levels of significance and power while minimizing the total expected budget. The Budget Minimization problem follows from a researcher's utility function and relies on two key ingredients. First, it relies on the power analysis to compute the required sample size for a predicted effect size. Second, it relies on a theoretical model to predict the outcomes in the treatment and control groups for a given level of incentives. The outcome of the Budget Minimization problem is the optimal level of incentives in the treatment group relative to the control, a variable I refer to as the *treatment strength*. The treatment strength pins down the required sample size, expected payoffs per subject, and the total expected budget.

The key tension in the Budget Minimization problem is between a required sample size and expected per-subject payoffs. On the one hand, increasing incentives leads to a higher expected effect size, which in turn drives down the required sample size and hence the expected total budget (the sample-size effect). On the other hand, increasing incentives leads to higher expected per-subject payoffs, which, in turn lead to a higher expected total budget (the payoff effect). My main theoretical result is that, under fairly mild assumptions, the Budget Minimization problem has a

(2015) (distributional choices), Parravano and Poulsen (2015) (symmetric coordination games), Yamagishi et al. (2016) (prisoner's dilemma), Schier, Ockenfels, and Hofmann (2016) (dictator game), Mengel (2017) (prisoner's dilemma), Bellemare, Sebald, and Suetens (2018) (psychological games), Leibbrandt and Lynham (2018) (common pool games).

[4] This is not always the case, e.g., researchers might wish to vary the level of the treatment variable to estimate a structural model (Andreoni and Miller, 2002; Holt and Laury, 2002; DellaVigna and Pope, 2018) or might be interested in the treatment effect of a specific value of a treatment variable. I do not consider these cases in what follows.

[5] Setting the appropriate level of a piece rate in performance pay experiments is notoriously difficult (Lazear, 2018; Carpenter and Huet-Vaughn, 2019).

non-trivial solution where the two effects are in the exact balance. I illustrate the properties of a solution using existing experiments, sketch a practical guide for setting up the problem and solving it in one's own design, and provide a sample R code.[6]

My main contribution is to offer a disciplined economic approach to the problem of choosing an optimal level of monetary incentives in experiments where incentives are a treatment variable. Experimental budgets are rarely explicitly discussed by researchers. Money, however, is a scarce resource, which makes it natural to ask what is an optimal way to use it. This question is of particular concern to junior scholars and PhD students, whose budgets are usually quite small while the pressure to produce significant results is high, as well as to researchers running expensive large-scale interventions in the field. This question is relevant both for new experiments[7] and replications.[8]

The logic of the Budget Minimization problem is quite general and extends beyond monetary incentives as a treatment variable. The approach applies to any treatment variable as long as it "behaves like money:" the treatment variable should create a tension between the sample-size effect and the payoff effect. The Budget Minimization problem is an alternative approach to an optimal experimental design that expands experimental economists' toolkit. The main point of departure from the traditional approach to an optimal design is the explicit inclusion of budget considerations. As an alternative approach, the Budget Minimization problem challenges some of the received wisdom in experimental design. For example, a common statistical, as well as design, recommendation is to set the values of a treatment variable as far apart as possible to ensure a maximum separation between predictions or a maximum variation in the treatment (Friedman and Sunder, 1994; List, Sadoff, and Wagner, 2011; Holt, 2019).[9] My approach shows that it may not be optimal to do this if separating the treatment values as much as possible leads to prohibitively high payoffs. Maximizing treatment strength, in other words, is not equivalent to maximizing a researcher's utility.

[6] The code is at https://github.com/aalexee/power_incentives.

[7] Even if a study is not a replication per se, it is common to replicate existing findings to establish a baseline before introducing a new treatment.

[8] An important qualification is that the replication will necessarily be conceptual, rather than direct (Camerer, Dreber, and Johannesson, 2019), in this case since the Budget Minimization problem will likely yield treatment values that are different from the ones in an original study.

[9] While this is true in many cases, there are some important exceptions, such as non-linear models (Moffatt, 2015).

2 Related Literature

The design of incentives has always been central to the methodology of experimental economics. Recent years, in particular, have seen a surge of interest in the theoretical analysis of this issue (Cox, Sadiraj, and Schmidt, 2014; Wilcox, 2018; Azrieli, Chambers, and Healy, 2018, 2020; Li, 2021; Johnson et al., 2021). Incentive compatibility of payoff mechanisms, or *how* to pay subjects, has so far dominated the analysis. The present work focuses instead on the optimal level of incentives, or *how much* to pay subjects, in case when incentives are a treatment variable.

The literature on optimal experimental designs is voluminous. Here I focus only on a few most closely related papers. In experimental economics, List, Sadoff, and Wagner (2011) and Vasilaky and Brock (2020) are recent concise guides to designing experiments and power analysis, while Moffatt (2015) and Holt (2019) are more comprehensive textbook treatments. The classic optimal experimental design literature deals primarily with the question of choosing the optimal number of subjects using the power analysis. More recent studies shift the focus to adaptive, or sequential, designs (Imai and Camerer, 2018; Balietti, Klein, and Riedl, 2021; Kasy and Sautmann, 2021; Johnson et al., 2021). A common theme among the more recent and classic works is that they take a statistical approach to the problem: the goal is to minimize the number of subjects or to maximize power, or to minimize the standard errors of an estimator. The power formula, in other words, is the objective function. In the present paper, I complement the statistical approach with the economic one. I make the power formula an input to a more general objective function—a researcher's utility function—that explicitly includes a budget.[10]

Several papers exploit structural models to guide experimental design. They typically use stochastic discrete-choice models (Moffatt, 2007; Rutström and Wilcox, 2009; Woods, 2021). The approach I take is similar to these works in that I also advocate for, and show the benefits of, using structural models for experimental design. The main difference, however, is that I make budget considerations explicitly a part of the optimization problem.

[10] List, Sadoff, and Wagner (2011) does discuss some cost considerations in designing an experiment, however, it does not explicitly make costs a part of an optimization problem.

3 Budget Minimization Problem

Consider a researcher planning a budget for an experiment. The expected total experimental budget, π, depends on the number of subjects in the experiment and expected per-subject payoffs. The researcher plans to use a standard between-subject design with two groups: control (C) and treatment (T). Let $G = \{C, T\}$ denote the set of experimental groups and $g \in G$ be its generic element. The researcher plans to use the same number of participants, n, in each group.[11] The researcher uses a single treatment variable. I assume that the treatment variable is monetary incentives.[12] Depending on the nature of the choice variable in the experiment, the researcher could use the difference in means or the difference in proportions as the outcome of interest.

Let τ_g denote the value of the treatment variable in group g. I denote the difference between the values of the treatment variable in the treatment and control groups as $\tau \equiv \tau_T - \tau_C, \tau \in \mathbb{R}_+$ and refer to it as the treatment strength. In some cases it can be of interest to have the treatment strength as a multiplicative factor rather than a difference. The above definition of the treatment strength accommodates these cases by defining the values of the treatment variable on a logarithmic scale. If one defines $\tau \equiv \ln \tilde{\tau}, \tau_g \equiv \ln \tilde{\tau}_g$, then $\tilde{\tau} = \exp(\tau) = \exp(\tau_T)/\exp(\tau_C)$ is the multiplicative treatment strength. I assume that the treatment strength is the only lever the researcher uses to optimize the budget.[13]

The researcher uses the power analysis to determine the required number of subjects in each group. This number will depend on the statistical parameters (significance α and power $1 - \beta$) and on the expected outcomes in each group, μ_g. The researcher sets significance and power at some conventional levels.[14] The expected outcomes can be, e.g., the mean choices in each group in case the choice variable is continuous or the proportions of subjects choosing a given alternative in case the outcome is discrete.[15] To predict the expected outcomes, the researcher uses a theoretical model parametrized by a vector of behavioral parameters γ. These parameters can include, e.g.,

[11]This assumption is not required and made to simplify the exposition.

[12]The Budget Minimization problem applies to non-monetary treatment variables, too, as long as they satisfy certain conditions. I discuss these conditions in Section 6.

[13]The researcher can exploit other design parameters to optimize the budget. However, those parameters are likely to be specific to each experiment. Hence it would difficult to obtain general results in that case.

[14]While relying on standards of significance thresholds is commonplace, the practice is not without issues (Brodeur, Cook, and Heyes, 2020; Brodeur et al., 2016).

[15]To be precise, I am calling a choice variable continuous if in the theoretical model it is a continuous function of the treatment variable, and the experiment allows subjects to make their choices among a large set of alternatives.

risk aversion, time preferences parameters, social preferences parameters, the curvature of the cost-of-effort function, etc. The researcher takes the behavioral parameters as given based on prior estimates. The expected outcomes will then depend on the treatment strength, behavioral parameters, as well as any other potential parameters of the experiment lumped in a vector δ: $\mu_g = \mu_g(\tau \mid \gamma, \delta)$. Vector δ includes things that are not explicitly modeled but that can nevertheless affect behavior, e.g., subject pool, number of rounds, framing of the instructions, whether a study is done in the lab or in the field, etc. To make everything a function of τ only, I use a convention that the level of incentives in the control group, τ_C, is included in vector δ. To summarize, the required number of subjects in each group depends on the parameters as follows: $n = n(\tau \mid \alpha, \beta, \gamma, \delta)$. It is worth emphasizing that the researcher does *not* pick n by itself, as is the case in a typical power analysis. Instead, she picks τ that affects expected outcomes that in turn pin down n, conditional on other parameters.

The expected per-subject payoffs in each group, π_g, will depend on expected outcomes and on the way the outcomes are translated into payoffs. I use a convention that these payoffs do not include the participation payment w. For example, when the outcome is the mean number of problems solved in a real-effort task and the treatment variable is a piece rate, the relationship between outcomes and payoffs takes a separable form: $\pi_g(\tau \mid \gamma, \delta) = \tau_g \mu_g(\tau \mid \gamma, \delta)$.

Assume that the researcher derives utility u from finding a true effect and 0 in all other cases.[16] The probability of finding a true effect is $1 - \beta$. The researcher's expected utility function from conducting the experiment is

$$U(\tau \mid \alpha, \beta, \gamma, \delta) = (1 - \beta)u - \pi(\tau \mid \alpha, \beta, \gamma, \delta). \tag{1}$$

Maximizing the researcher's utility function is equivalent to minimizing the Budget Minimization problem.[17]

$$\min_{\tau} \pi(\tau \mid \alpha, \beta, \gamma, \delta) = n(\tau \mid \alpha, \beta, \gamma, \delta) \left(2w + \pi_C(\gamma, \delta) + \pi_T(\tau \mid \gamma, \delta)\right). \tag{2}$$

[16]The implicit assumption here is that the researcher is not nefarious and does not derive utility from finding a false positive.

[17]I formulate this problem without any constraints for simplicity. I discuss constraints in Section 6.

The intuition for why the Budget Minimization problem makes sense is the following. The response of the budget to a change in the treatment strength depends on two effects: the sample-size effect and the payoff effect. Increasing τ is expected to increase the difference in outcomes between the treatment and control groups. The predicted effect size will increase, which in turn will drive down the required number of subjects (the sample-size effect). On the other hand, increasing τ will increase the expected per-subject payoff in the treatment group due to the direct effect of higher incentives and the indirect effect of higher outcomes due to higher incentives (the payoff effect). These two opposing effects can potentially lead to a point τ^* where the expected total budget is minimized.

Formally, the following first-order necessary condition must hold at the optimal point τ^*[18]

$$-\frac{n'(\tau)}{n(\tau)} = \frac{\pi_T'(\tau)}{2w + \pi_C(\tau) + \pi_T(\tau)}. \tag{3}$$

The condition states, intuitively, that at the optimum the percentage decrease in the required number of subjects due to the higher treatment strength (the sample-size effect) exactly offsets the percentage increase in the per-subject payoffs (the payoff effect). The theoretical question is under what conditions the Budget Minimization problem has a non-trivial solution. Before I turn to the formal analysis of this question, I present two examples of the Budget Minimization problem at work.

4 Budget Minimization in Practice

I illustrate the Budget Minimization problem in two common cases. In the first case, the choice variable is continuous and the outcome of interest is the difference in mean choices. In the second case, the choice variable is discrete. In this case, the outcome of interest can be either the difference in proportions of subjects choosing a given alternative (binary choice) or the difference in mean choices (more than two alternatives). I focus on the former case when the choice is binary and the outcome of interest is the difference in proportions, although a similar logic would apply to the latter case.

[18] To avoid notational clutter, I drop the dependence on the parameters $\alpha, \beta, \gamma, \delta$.

4.1 Continuous Case

To illustrate the Budget Minimization problem in the continuous case, I use the experiment of DellaVigna and Pope (2018). In the experiment, subjects perform a real-effort task in which they have to repeatedly press two buttons for ten minutes. Subjects receive $w = \$1$ for their participation. The choice variable is the number of button presses, a proxy for a subject's effort. The outcome variable is the average number of button presses.

Suppose that the researcher is interested in testing whether introducing a piece rate in the treatment group increases effort relative to the control group that receives no piece rate, $\tau_C = 0$. The expected per-subject payoff in group g is $\pi_g = \tau_g \mu_g$. Subjects receive a piece rate for each 100 button presses. The goal is to determine the treatment strength τ that allows one to detect an increase in effort for the conventional levels of significance ($\alpha = 0.05$) and power ($1 - \beta = 0.8$) while minimizing the required budget.

DellaVigna and Pope (2018)[P. 1063] propose a model of effort choice that gives the following closed-form solution for the mean effort:[19]

$$\mu_g(\tau \mid \gamma, \delta) = \frac{1}{\eta}[\ln(s + \tau_g) - \ln k]. \tag{4}$$

where η and k are the curvature and scale parameters of the cost-of-effort function, respectively, s is an intrinsic reward for performing the task, and τ_g is a piece rate in group g. In the notation of the Budget Minimization problem, the vector of behavioral parameters γ is then (η, k, s, σ).

Knowing the formula for the expected outcomes, one can find the required number of subjects per group conditional on τ and other parameters for a two-sided t-test for the difference in means:[20]

$$n(\tau \mid \alpha, \beta, \gamma, \delta) = 2\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \left(\frac{\sigma}{\mu_T(\tau \mid \gamma, \delta) - \mu_C(\gamma, \delta)}\right)^2. \tag{5}$$

Then, using formula (2) one can find the expected total budget as a function of τ.

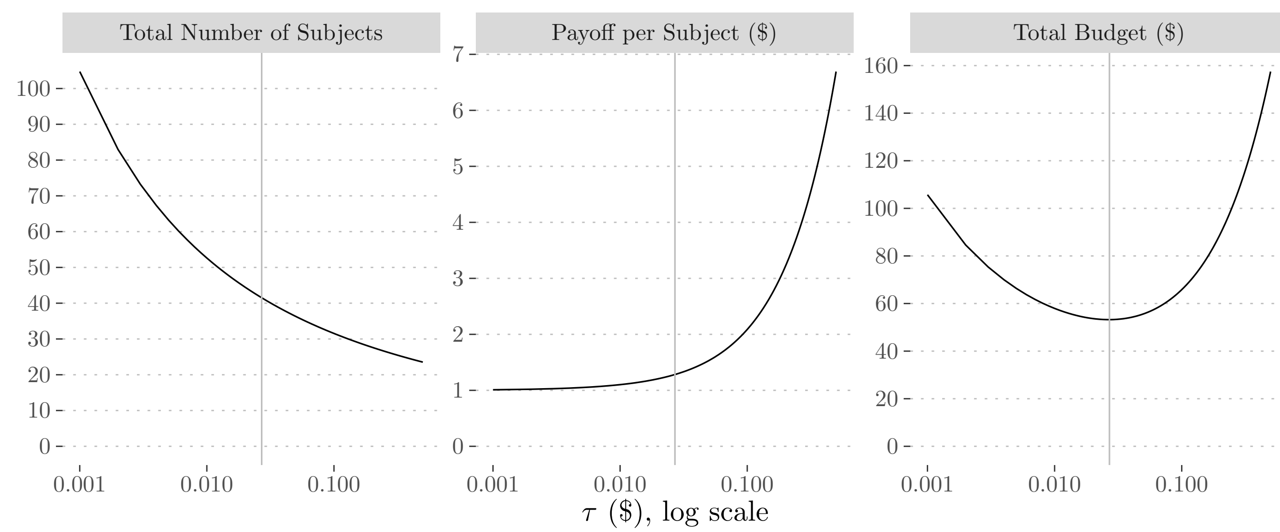
Figure 1 shows how the variables of the experiment change with τ. The total number of

[19] Specifically, I use the version of their model with the exponential cost of effort. I make several changes to the authors' original notation to make it consistent with the notation adopted in my paper. In their formula (13), I substitute γ for η and p for τ_g.

[20] I use the following parameter estimates, $\eta = 0.015641071, k = 1.70926702 \times 10^{-16}, s = 3.72225938 \times 10^{-6}, \sigma = 653.578104$ (Supplementary Material "NLS_results_Table5_EXPON.csv:").

subjects across both groups, $2n$, decreases in τ since higher incentives increase the expected effect size. The expected payoff per subject across both groups, $w + (\pi_C + \pi_T)/2$, increases in τ since higher incentives increase expected effort, as well as the payoff per unit of effort. The expected total budget π reaches a minimum at $\tau^* = 3$ cents (the result is rounded to the nearest cent). Conducting such an experiment would be extremely cheap: the experiment would require a total of 42 subjects with an expected per-subject payoff across both groups of \$1.28 and an expected total budget of just \$53.[21] For comparison, the original experiment has 0 and 4 cents treatments, although the total number of subjects in both groups is more than 1000. Changing the statistical significance to $\alpha = 0.001$ and power to $1 - \beta = 0.99$ would still be highly affordable: it would require a total sample size of 167 subjects and an expected total budget of \$214.[22]

Figure 1: Variables of the DellaVigna and Pope (2018) Experiment as a Function of τ



Note: The figure shows how the three variables of the experiment change with the treatment strength τ. The left panel shows the total number of subjects across both treatment groups ($2n$). The middle panel shows the expected per-subject payoff (in \$) across both treatment groups ($w + (\pi_C + \pi_T)/2$). The right panel shows the expected total budget (in \$) (π). The horizontal axis shows the treatment strength τ (in \$) on a logarithmic scale. The vertical solid line shows the budget-minimizing level of τ.

[21] While these numbers do appear small, they are not unreasonable given the large treatment effects found in the data. For instance, the mean effort levels in the 0 and 4 cents treatments are 1521 and 2132, respectively, (DellaVigna and Pope, 2018)[P. 1045, Table 3]. Assuming a common standard deviation of 650, the traditional power analysis would yield 18 subjects per treatment group for the levels of significance (0.05) and power (0.8) assumed in my calculation. Running an experiment with so few subjects, certainly, would not be advisable. The calculations made here are done without any constraints on n or per-subject payoffs.

[22] Note that changing α or β only affects n and π, but does not affect the optimal level of incentives or per-subject payoffs.

The results, however, turn out to be highly sensitive to the level of incentives in the control group. Suppose that the researcher wishes to test the hypothesis that increasing the piece rate from $\tau_C = 1$ cent can increase effort. Setting $\alpha = 0.05$ and $1 - \beta = 0.8$, the optimal level of incentives in the treatment group would be 19 cents. The experiment would require a total of 376 subjects with an expected per-subject payoff across both groups of \$3.23 and an expected total budget of \$1213. The reason for such a steep increase in the budget is that raising the control group incentives dramatically increases the expected effort μ_C.

4.2 Discrete Choice

To illustrate the Budget Minimization problem in the discrete-choice setting, I use the classic Holt and Laury (2002) experiment on risk aversion. In this experiment, which popularized the multiple-price-list elicitation method, subjects make a series of binary choices between a safe and risky lotteries. The alternatives are ordered such that a risky lottery gradually becomes more attractive. Experimental treatments involve changing the level of incentives by large factors to see whether this affects the proportion of subjects choosing a safe lottery.

For illustrative purposes, suppose that the researcher is interested in testing whether scaling the payoffs of each lottery up affects the proportion of subjects choosing a safe lottery in just one pair.[23] Suppose the researcher picks pair 5 (Holt and Laury, 2002)[P. 1645, Table 1] in which the safe lottery pays \$2 or \$1.6 with equal chances and the risky lottery pays \$3.85 or \$0.1 with equal chances in the control group, and in which the safe lottery pays \$2$\times\tau$ or \$1.6$\times\tau$ with equal chances and the risky lottery pays \$3.85$\times\tau$ or \$0.1$\times\tau$ with equal chances in the treatment group. Here τ is the multiplicative treatment strength. The expected per-subject payoff in group g is $\pi_g = \tau_g(\mu_g EV_A + (1 - \mu_g)EV_B)$, where EV_A and EV_B are the expected values of the safe and risky lotteries, respectively, in the control group ($\tau_C = 1$) and μ_g is the proportion of subjects choosing the safe lottery in group g. The goal is to determine the treatment strength that allows the researcher to detect a change in the proportion of subjects choosing the safe lottery for the conventional levels of significance ($\alpha = 0.05$) and power ($1-\beta = 0.8$) while minimizing the required

[23] Discrete choice does not necessarily imply that the relevant outcome is the proportion of subjects choosing a given alternative. While it is true in the binary choice, in case when there are more than two alternatives a researcher might consider the difference in mean choices. In the context of Holt and Laury (2002), this could be, e.g., the mean switching point. The models of stochastic discrete choice, such as the one considered here, can still be used to derive the expected outcomes in the case of more than two alternatives.

budget.

Holt and Laury (2002) use the stochastic choice model that specifies the probability of choosing the safe lottery in group g as follows:[24]

$$\mu_g(\tau \mid \gamma, \delta) \equiv \mathbb{P}(A)_g = \frac{U_{A_g}^{1/\lambda}}{U_{A_g}^{1/\lambda} + U_{B_g}^{1/\lambda}}, \tag{6}$$

where U_{A_g}, U_{B_g} are the expected utilities of the safe and risky lotteries, respectively, in group g and λ is the noise parameter. The expected utility uses an expo-power utility-of-money function of the form[25]

$$u(x) = \frac{1 - \exp(-ax^{1-r})}{a}, \tag{7}$$

where x is a monetary outcome, a is the constant risk aversion parameter, and r is the relative risk aversion parameter. The vector of behavioral parameters γ is then composed of (a, r, λ).

Knowing the formula for the probability of choosing the safe option, one can find the required number of subjects per group conditional on τ and other parameters for a test for the difference in proportions:[26]

$$n(\tau \mid \alpha, \beta, \gamma, \delta) = (z_{1-\alpha/2} + z_{1-\beta})^2 \frac{\mu_T(1-\mu_T) + \mu_C(1-\mu_C)}{(\mu_T - \mu_C)^2}. \tag{8}$$

Then using formula (2) and the formulas for expected per-subject payoffs, one can find the expected total budget as a function of τ.[27]

Figure 2 shows how the variables of the experiment change with τ. The total number of subjects across both groups decreases in τ, while the expected payoff per subject across both groups increases in τ. The expected total budget reaches a minimum at $\tau^* = 54$ (rounded to the nearest digit). This means that the payoffs need to be scaled by more than 50 times. The experiment would require a total of 99 subjects with an expected per-subject payoff across both groups of \$55.1

[24] A more common choice for the stochastic model would have been the Fechner model (Wilcox, 2008). I stick to the original specification for comparability.

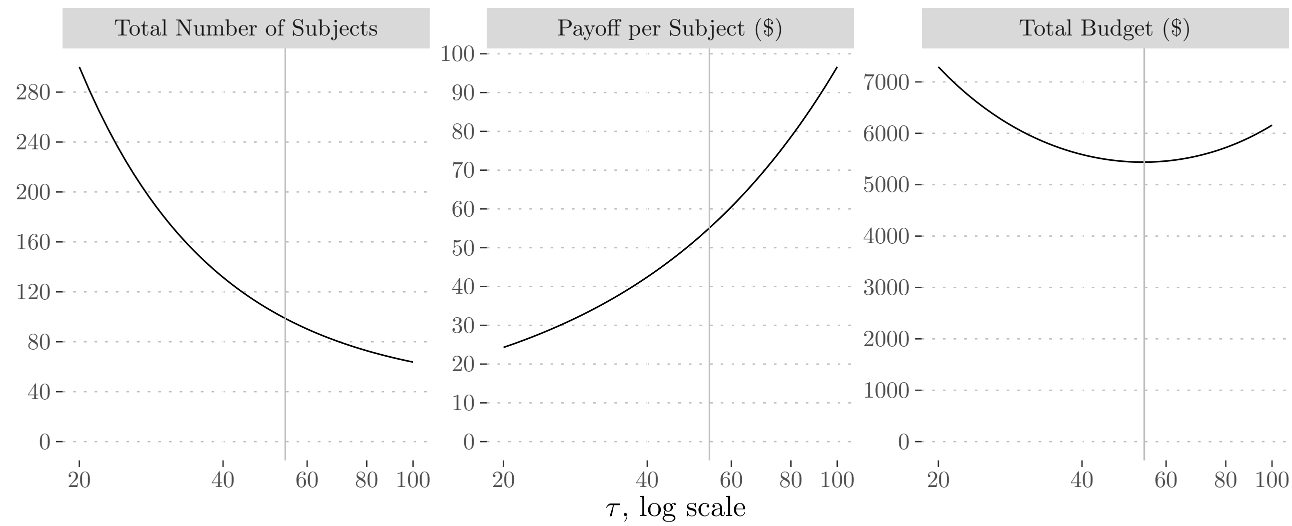
[25] I substitute α for a in the authors' original specification ((Holt and Laury, 2002)[P. 1653, formula (2)]) to avoid confusion with the significance level α. I also substitute μ for λ in formula (1).

[26] To avoid notational clutter, I drop the dependence of μ_T and μ_C on τ, γ, δ. I use the following parameter estimates (Holt and Laury, 2002)[P. 1653]: $a = 0.029, r = 0.269, \lambda = 0.134$.

[27] While the participation payment is not explicitly mentioned in the text, I assume $w = \$5$, which is a typical amount for laboratory experiments.

and an expected total budget of $5439. For comparison, the original experiment does have a 50x treatment, although the number of subjects in this group is only 19.

Figure 2: Variables of the Holt and Laury (2002) Experiment as a Function of τ



Note: The figure shows how the three variables of the experiment change with the treatment strength. The left panel shows the total number of subjects across both treatment groups ($2n$). The middle panel shows the expected per-subject payoff (in $) across both treatment groups ($w + (\pi_C + \pi_T)/2$). The right panel shows the expected total budget (in $) ($\pi$). The horizontal axis shows the multiplicative treatment strength τ on a logarithmic scale. The vertical solid line shows the budget-minimizing level of τ.

4.3 Practical Guide

To facilitate the use of the Budget Minimization problem in practice, I sketch a recipe for how to set it up and solve numerically. I provide the R code to reproduce the two examples above at https://github.com/aalexee/power_incentives.

- Define the outcome functions μ_C and $\mu_T(\tau)$. These functions will come from a theoretical model of choice.
- Use μ_C and $\mu_T(\tau)$ to define the sample size as a function of τ, $n(\tau)$, using the relevant power formula, e.g., (5) or (8).
- Use μ_C and $\mu_T(\tau)$ to define the payoff functions π_C and $\pi_T(\tau)$.

- Combine the functions $n(\tau)$, π_C, and $\pi_T(\tau)$ using formula (2) to get the function for the total budget $\pi(\tau)$.
- Minimize $\pi(\tau)$ numerically, denote τ^* the value that minimizes the budget. It usually helps to plot the total budget with the logarithm of τ on the x-axis.
- Use τ^* to compute the total required sample size $2n(\tau^*)$, expected per-subject payoff across both groups $w + \pi_C/2 + \pi_T(\tau^*)/2$, and total expected budget $\pi(\tau^*)$.

5 Budget Minimization in Theory

I make two assumptions about the outcome function $\mu_T(\tau)$ to establish a theoretical result.

Assumption 1 (Continuous Differentiability). $\mu_T \in C^1$.

Assumption 2 (Regularity). $\lim_{\tau \to \tau^{low}} |\mu'_T| < \infty$ *and* $\lim_{\tau \to \infty} d \ln \mu_T / d \ln \tau < 1$.[28]

The first assumption is a technical one. The second assumption takes care of the case when μ_T is unbounded. In this case, it has to satisfy regularity conditions that require the outcome function a) not to change too quickly when treatment increases from the lowest value and b) that the elasticity of the outcome with respect to τ is small as the treatment strength gets large. Assumption 2 is satisfied automatically if μ_T is bounded.

Proposition 1. *If μ_T satisfies Assumptions 1 and 2 the Budget Minimization problem has an interior solution.*

Proof. See Appendix A. □

The idea of the proof relies on the Intermediate Value Theorem and the properties of the two components of the total budget: the sample size and expected payoffs.[29] I consider the limiting behavior of the derivative of the logarithm of the total budget with respect to τ. At the lower limit, when the treatment strength approaches the lower bound, the derivative of the budget goes

[28] Here τ^{low} denotes the lowest possible value of τ. It is 0 for additive treatment strength and 1 for multiplicative treatment strength.

[29] One might wonder if the Weierstrass theorem would suffice instead. It would not: even if one is willing to impose an upper bound on τ (which is a priori unclear), the Weierstrass theorem cannot say anything about an interior solution, which is the interesting case.

to negative infinity. The driver behind this result is the required sample size. When the treatment strength is zero (additive case) or one (multiplicative case) the outcomes in the treatment and control groups are identical, which makes the required sample size infinite. Even the smallest increase in the treatment strength is enough to produce an infinitely large decrease in the required sample size. At the lower limit, therefore, the negative sample-size effect dominates the positive payoff effect.[30] When the treatment strength is infinitely large, neither the required sample size nor the expected payoffs change. The derivative of the total budget in the limit is zero. However, one can always find a large enough value of the treatment strength at which the derivative of the total budget is positive. At the upper limit, therefore, the positive payoff effect dominates the negative sample-size effect.[31] The derivative of the total budget is thus negative at the left endpoint and positive at the right endpoint. Since μ_T is continuously differentiable by Assumption 1, the Intermediate Value Theorem implies that the derivative of the total budget must cross zero. Since the first crossing will occur from below, the First Order Sufficient Condition for a Minimum implies that the point τ^* at which this happens must be a minimum point.

The result in Proposition 1 is surprisingly general. It applies both in the continuous and discrete cases. The assumptions required for the result are fairly weak. The discrete case effectively only requires Assumption 1, since the outcome is a proportion bounded between zero and one. The continuous case would in addition require Assumption 2 only if the outcome function is unbounded.

Proposition 1 explains why the motivating examples work. In the discrete case example, only Assumption 1 needs to be checked. Indeed, since the utility-of-money function (7) is continuously differentiable, so are the expected utility and outcome (6) functions. Proposition 1 immediately applies. In the continuous case example, the outcome function (4) is continuously differentiable but unbounded, hence we need to check Assumption 2, as well. First, consider

$$\lim_{\tau \to 0^+} |\mu_T'| = \lim_{\tau \to 0^+} \frac{1}{\eta(s+\tau)} = \frac{1}{\eta s}.$$

[30]If the outcome function is unbounded, the first part of Assumption 2 guarantees that.

[31]If the outcome function is unbounded, the second part of Assumption 2 guarantees that.

The limit is finite, since the estimates of s and η are strictly positive. On the other hand,

$$\lim_{\tau\to\infty} \tau(\ln\mu_T)' = \lim_{\tau\to\infty} \frac{\tau}{(s+\tau)\ln\left(\frac{s+\tau}{k}\right)} = \lim_{\tau\to\infty} \frac{1}{(\frac{s}{\tau}+1)\ln\left(\frac{s+\tau}{k}\right)} = 0 < 1,$$

provided that $k > 0$, which is indeed the case given the model estimates. Hence, Proposition 1 also applies.

A few remarks about the theoretical result are in order. The first remark is that Assumptions 1 and 2 are sufficient but not necessary. It might as well be that they are not satisfied but the Budget Minimization problem has an interior solution. The second, and related, remark is that Assumption 1 can have a bite in some cases. It might fail to hold in reference-dependent models, which feature a discontinuity around a reference point. The budget, however, is still likely to have a minimum. The third, and final, remark is that Proposition 1 guarantees the existence but not the uniqueness of a solution. It is safe to assume that it should not cause any issues in practice. If there are several minimum points, one can simply compute the budget at each of the candidate solutions and pick the one giving the smallest budget.

6 Discussion

In this section, I propose some extensions of the Budget Minimization problem and show that its applicability goes beyond the examples analyzed so far. I also highlight some of the limits of its applicability.

Non-Monetary Treatment

The working assumption in setting up the Budget Minimization problem has been that the treatment variable is money. The structure of the problem, however, does not require the treatment variable to be money, it just has to "behave like money." First, the treatment variable has to be finely divisible. If the treatment variable can assume only a few possible values (e.g., the communication between subjects is either allowed or not), the question about optimizing the value of the treatment variable is meaningless. Second, increasing the treatment variable has to increase the expected payoff in the treatment group. If this condition does not hold, there is no tension between the sample-size effect and the payoff effect. The solution to the Budget Minimization problem would be trivial: simply pick the highest possible value of τ. Those are the necessary conditions for the Budget Minimization problem to be meaningful. The sufficient conditions will depend on the exact way that the treatment variable and the outcome function μ_T affect the payoff π_T.

Strategic Settings

Even though the examples I considered are from individual-choice settings, the logic of the Budget Minimization problem carries over to strategic settings. The natural counterpart to the theoretical outcome function μ_T, such as (6), in game theory is the Quantal Response Function (McKelvey and Palfrey, 1995; Goeree, Holt, and Palfrey, 2005).

Parameter Uncertainty

The solution to the Budget Minimization problem relies on the estimates of the structural parameters of a model. These estimates will have standard errors. The analysis conducted in motivating examples ignores this parameter uncertainty for simplicity. However, the budget-minimizing treatment strength is a function of parameters and hence inherits the uncertainty in their estimates. The optimal treatment strength is unlikely to have a closed-form solution in most cases, hence, using the Delta method would be impossible. A practical solution to deriving the standard errors of the treatment strength would be to use the bootstrap.

Parameter Estimates

A related, and more fundamental, point about parameter estimates is that they have to exist in order to take advantage of the Budget Minimization problem.[32] In the best-case scenario, these estimates could be readily available from the literature. This is likely to be the case for the models of risk and time preferences (Harrison and Rutström, 2008), lying aversion (Abeler, Nosenzo, and Raymond, 2019), social preferences (Goeree, Holt, and Laury, 2002; Cox, Friedman, and Gjerstad, 2007; Bellemare, Kröger, and Van Soest, 2008), and real-effort tasks (DellaVigna and Pope, 2018). But what should a researcher do when those estimates are not available or cannot be used? One

[32] This is an issue not just for the Budget Minimization problem but for optimal experimental design in general (List, Sadoff, and Wagner, 2011; Moffatt, 2015).

possibility is that a researcher can use an existing structural model but does not want to use existing parameter estimates. Using existing estimates might not be reliable if, e.g., they are derived from a subject pool that is very different from a researcher's subject pool. In other words, a researcher might worry about the portability of the existing estimates. A solution in this case is to run pilot sessions on the subject pool of interest and estimate the parameters of the model using the pilot data. Using pilots to conduct the power analysis is a standard practice in experimental economics, and the only modification to that practice would be the way the data are used. Another possibility is that an off-the-shelf structural model simply does not exist. In this case a researcher is left with the option to come up with their own model to take advantage of the Budget Minimization problem.

Constraints

I have presented and analyzed the Budget Minimization problem as an unconstrained problem. In reality, a researcher might face constraints on subjects' payoffs and/or a sample size. Suppose the sample size at τ^* is too low to be acceptable (the constraint binds), as we saw in the continuous case example. A researcher can simply tweak the statistical parameters: decreasing α or β increases the optimal sample size without changing the optimal treatment strength. Suppose now that the expected per-subject payoffs are too low at τ^*. In this case the optimal treatment strength will have to change. There are several possibilities to satisfy the constraint in that case. One possibility is to change the level of the treatment variable in the control group, re-optimize, and check if the constraint is satisfied. As we have seen in the continuous case example, adjusting τ_C can have a large impact on τ^*. The benefit of this approach is that one can both satisfy the constraint and get an optimal level of τ. Another possibility is to keep increasing τ until the constraint is satisfied. This approach will distort τ away from the budget-minimizing level. However, it can be more cost-effective than increasing τ_C. One might also consider changing the participation payment w, which will change τ^*. The participation payment, however, is typically set by lab policies and rarely tweaked for the purposes of a particular experiment.[33] On the other end of the spectrum is the case when the expected per-subject payoffs are too high. No simple solution exists in this case,

[33] A notable exception is when the participation payment *is* the treatment variable (Harrison, Lau, and Rutström, 2009).

since τ^* already minimizes the budget and any deviation will only increase it. A researcher would likely have to reconsider other parameters of the design to bring down the budget.

Non-Parametric Tests

In practice, researchers often use non-parametric tests, such as the Wilcoxon-Mann-Whitney test, to analyze treatment effects. The reason for relying on parametric tests in my analysis is that they have simple analytical formulas for power calculations and require only minimal predictions about outcomes, such as averages. Power analysis for non-parametric tests, on the other hand, is based either on simulations (Bellemare, Bissonnette, and Kröger, 2016) in which case deriving theoretical results is impossible, or on explicit formulas that require rich predictions about outcomes, such as the entire distribution of outcomes (Rahardja, Zhao, and Qu, 2009; Happ, Bathke, and Brunner, 2019). One can still pose a practical question about the optimal level of incentives for a non-parametric test in a given experiment and use simulations to solve the Budget Minimization problem.

7 Conclusion

I study an optimal design of incentives in experiments where incentives are a treatment variable. Using a utility-based framework, I formulate a Budget Minimization problem. In the problem, a researcher chooses a treatment strength such that it minimizes the expected budget while allowing to find an effect for the given levels of statistical significance and power. The effect of the treatment strength on the budget can be decomposed into two channels: the sample-size effect and the payoff effect. Increasing the treatment strength decreases the required budget via the sample-size effect but increases it via the payoff effect. At a minimum point, the two effects must be in the exact balance. I show theoretically that such a point exists under fairly mild conditions, and thus the Budget Minimization problem is guaranteed to have a non-trivial solution. I illustrate how the Budget Minimization problem applies in practice using existing experiments. The Budget Minimization problem also applies, under certain conditions, to designs where a treatment variable is not monetary incentives.

The main challenge in taking advantage of my approach is having a structural model and reliable prior estimates of the model, in other words, good prior data, albeit this is true in general for any optimal design. The main contribution of my analysis is that it takes the guesswork out of the design of the level of incentives and replaces it with a disciplined economic approach. I believe that my approach to the design of incentives will enrich experimental economists' toolkit and help guide future designs. Young researchers on tight budgets and researchers running expensive field interventions will particularly benefit from using the Budget Minimization problem.

References

Abeler, J., D. Nosenzo, and C. Raymond. 2019. "Preferences for Truth-Telling." *Econometrica* 87:1115–1153.

Amir, O., D.G. Rand, and Y.K. Gal. 2012. "Economic Games on the Internet: The Effect of $1 Stakes." *PloS one* 7:e31461.

Andreoni, J., and J. Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica* 70:737–753.

Araujo, F.A., E. Carbone, L. Conell-Price, M.W. Dunietz, A. Jaroszewicz, R. Landsman, D. Lamé, L. Vesterlund, S.W. Wang, and A.J. Wilson. 2016. "The Slider Task: An Example of Restricted Inference on Incentive Effects." *Journal of the Economic Science Association* 2:1–12.

Azrieli, Y., C.P. Chambers, and P.J. Healy. 2018. "Incentives in Experiments: A Theoretical Analysis." *Journal of Political Economy* 126:1472–1503.

—. 2020. "Incentives in Experiments with Objective Lotteries." *Experimental Economics* 23:1–29.

Balasubramanian, P., V.M. Bennett, and L. Pierce. 2017. "The Wages of Dishonesty: The Supply of Cheating Under High-Powered Incentives." *Journal of Economic Behavior & Organization* 137:428–444.

Balietti, S., B. Klein, and C. Riedl. 2021. "Optimal Design of Experiments to Identify Latent Behavioral Types." *Experimental Economics* 24:772–799.

Bellemare, C., L. Bissonnette, and S. Kröger. 2016. "Simulating Power of Economic Experiments: The powerBBK Package." *Journal of the Economic Science Association* 2:157–168.

Bellemare, C., S. Kröger, and A. Van Soest. 2008. "Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities." *Econometrica* 76:815–839.

Bellemare, C., A. Sebald, and S. Suetens. 2018. "Heterogeneous Guilt Sensitivities and Incentive Effects." *Experimental Economics* 21:316–336.

Brañas-Garza, P., P. Kujal, and B. Lenkei. 2019. "Cognitive Reflection Test: Whom, How, When." *Journal of Behavioral and Experimental Economics* 82:101455.

Brodeur, A., N. Cook, and A. Heyes. 2020. "Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics." *American Economic Review* 110:3634–60.

Brodeur, A., M. Lé, M. Sangnier, and Y. Zylberberg. 2016. "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics* 8:1–32.

Camerer, C.F., A. Dreber, and M. Johannesson. 2019. "Replication and Other Practices for Improving Scientific Quality in Experimental Economics." In A. Schram and A. Ule, eds. *Handbook of Research Methods and Applications in Experimental Economics*. Cheltenham, UK: Edward Elgar Publishing, pp. 83–102.

Camerer, C.F., and R.M. Hogarth. 1999. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework." *Journal of Risk and Uncertainty* 19:7–42.

Carpenter, J., and E. Huet-Vaughn. 2019. "Real-Effort Tasks." In A. Schram and A. Ule, eds. *Handbook of Research Methods and Applications in Experimental Economics*. Cheltenham, UK: Edward Elgar Publishing, pp. 368–383.

Cloos, J., M. Greiff, and H. Rusch. 2020. "Editorial Favoritism in the Field of Laboratory Experimental Economics." Research Memoranda No. 005, GSBE.

Cox, J.C., D. Friedman, and S. Gjerstad. 2007. "A Tractable Model of Reciprocity and Fairness." *Games and Economic Behavior* 59:17–45.

Cox, J.C., and D.M. Grether. 1996. "The Preference Reversal Phenomenon: Response Mode, Markets and Incentives." *Economic Theory* 7:381–405.

Cox, J.C., and V. Sadiraj. 2019. "Incentives." In A. Schram and A. Ule, eds. *Handbook of Research Methods and Applications in Experimental Economics*. Cheltenham, UK: Edward Elgar Publishing, pp. 9–27.

Cox, J.C., V. Sadiraj, and U. Schmidt. 2014. "Paradoxes and Mechanisms for Choice under Risk." *Experimental Economics*, pp. 1–36.

DellaVigna, S., and D. Pope. 2018. "What Motivates Effort? Evidence and Expert Forecasts." *The Review of Economic Studies* 85:1029–1069.

El Harbi, S., I. Bekir, G. Grolleau, and A. Sutan. 2015. "Efficiency, Equality, Positionality: What Do People Maximize? Experimental vs. Hypothetical Evidence from Tunisia." *Journal of Economic Psychology* 47:77–84.

Enke, B., U. Gneezy, B. Hall, D.C. Martin, V. Nelidov, T. Offerman, and J. van de Ven. 2021. "Cognitive Biases: Mistakes or Missing Stakes?" Working Paper No. 28650, National Bureau of Economic Research, April.

Fischbacher, U., and F. Föllmi-Heusi. 2013. "Lies in Disguise—An Experimental Study on Cheating." *Journal of the European Economic Association* 11:525–547.

Friedman, D., and S. Sunder. 1994. *Experimental Methods: A Primer for Economists*. Cambridge University Press.

Gibson, R., C. Tanner, and A.F. Wagner. 2013. "Preferences for Truthfulness: Heterogeneity among and within Individuals." *American Economic Review* 103:532–48.

Gneezy, U., S. Meier, and P. Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." *Journal of Economic Perspectives* 25(4):191–210.

Goeree, J.K., C.A. Holt, and S.K. Laury. 2002. "Private Costs and Public Benefits: Unraveling the Effects of Altruism and Noisy Behavior." *Journal of Public Economics* 83:255–276.

Goeree, J.K., C.A. Holt, and T.R. Palfrey. 2005. "Regular Quantal Response Equilibrium." *Experimental Economics* 8:347–367.

Grether, D.M., and C.R. Plott. 1979. "Economic Theory of Choice and the Preference Reversal Phenomenon." *American Economic Review* 69:623–638.

Happ, M., A.C. Bathke, and E. Brunner. 2019. "Optimal Sample Size Planning for the Wilcoxon-Mann-Whitney Test." *Statistics in Medicine* 38:363–375.

Harrison, G.W., M.I. Lau, and E.E. Rutström. 2009. "Risk Attitudes, Randomization to Treatment, and Self-Selection into Experiments." *Journal of Economic Behavior & Organization* 70:498–507.

Harrison, G.W., and E.E. Rutström. 2008. "Risk Aversion in the Laboratory." In J. C. Cox and G. W. Harrison, eds. *Risk Aversion in Experiments (Research in Experimental Economics, Vol. 12)*. Emerald Group Publishing Limited, Bingley, pp. 41–196.

Hertwig, R., and A. Ortmann. 2001. "Experimental Practices in Economics: A Methodological Challenge for Psychologists?" *Behavioral and Brain Sciences* 24:383–403.

Holt, C.A. 2019. *Markets, Games, and Strategic Behavior: An Introduction to Experimental Economics*. Princeton University Press.

Holt, C.A., and S.K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92:1644–1655.

Imai, T., and C.F. Camerer. 2018. "Estimating Time Preferences from Budget Set Choices Using Optimal Adaptive Design." Working paper, LMU Munich.

Johnson, C., A. Baillon, H. Bleichrodt, Z. Li, D. van Dolder, and P.P. Wakker. 2021. "Prince: An Improved Method for Measuring Incentivized Preferences." *Journal of Risk and Uncertainty* 62:1–28.

Kahneman, D., and A. Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, pp. 263–291.

Kasy, M., and A. Sautmann. 2021. "Adaptive Treatment Assignment in Experiments for Policy Choice." *Econometrica* 89:113–132.

Kleinlercher, D., and T. Stöckl. 2018. "On the Provision of Incentives in Finance Experiments." *Experimental Economics* 21:154–179.

Larney, A., A. Rotella, and P. Barclay. 2019. "Stake Size Effects in Ultimatum Game and Dictator Game Offers: A Meta-Analysis." *Organizational Behavior and Human Decision Processes* 151:61–72.

Lazear, E.P. 2018. "Compensation and Incentives in the Workplace." *Journal of Economic Perspectives* 32(3):195–214.

Leibbrandt, A., and J. Lynham. 2018. "Does the Paradox of Plenty Exist? Experimental Evidence on the Curse of Resource Abundance." *Experimental Economics* 21:337–354.

Li, Y. 2021. "The ABC Mechanism: An Incentive Compatible Payoff Mechanism for Elicitation of Outcome and Probability Transformations." *Experimental Economics* 24:1019–1046.

List, J.A., S. Sadoff, and M. Wagner. 2011. "So You Want to Run an Experiment, Now What? Some Simple Rules of Thumb for Optimal Experimental Design." *Experimental Economics* 14:439.

McKelvey, R.D., and T.R. Palfrey. 1995. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior* 10:6–38.

Mengel, F. 2017. "Risk and Temptation: A Meta-study on Prisoner's Dilemma Games." *The Economic Journal* 128:3182–3209.

Moffatt, P.G. 2015. *Experimetrics: Econometrics for Experimental Economics*. Palgrave Macmillan.

—. 2007. "Optimal Experimental Design in Models of Decision and Choice." In M. Boumans, ed. *Measurement in Economics: A Handbook*. London: Academic Press, pp. 357–377.

Parravano, M., and O. Poulsen. 2015. "Stake Size and the Power of Focal Points in Coordination Games: Experimental Evidence." *Games and Economic Behavior* 94:191–199.

Pulford, B.D., A.M. Colman, and G. Loomes. 2018. "Incentive Magnitude Effects in Experimental Games: Bigger is not Necessarily Better." *Games* 9.

Rahardja, D., Y.D. Zhao, and Y. Qu. 2009. "Sample Size Determinations for the Wilcoxon–Mann–Whitney Test: A Comprehensive Review." *Statistics in Biopharmaceutical Research* 1:317–322.

Rousu, M.C., J.R. Corrigan, D. Harris, J.K. Hayter, S. Houser, B.A. Lafrancois, O. Onafowora, G. Colson, and A. Hoffer. 2015. "Do Monetary Incentives Matter in Classroom Experiments? Effects on Course Performance." *The Journal of Economic Education* 46:341–349.

Rutström, E.E., and N.T. Wilcox. 2009. "Stated Beliefs versus Inferred Beliefs: A Methodological Inquiry and Experimental Test." *Games and Economic Behavior* 67:616–632.

Schier, U.K., A. Ockenfels, and W. Hofmann. 2016. "Moral Values and Increasing Stakes in a Dictator Game." *Journal of Economic Psychology* 56:107–115.

Smith, V.L. 1976. "Experimental Economics: Induced Value Theory." *American Economic Review* 66:274–279.

—. 1982. "Microeconomic Systems as an Experimental Science." *American Economic Review* 72:923–955.

Smith, V.L., and J.M. Walker. 1993. "Rewards, Experience And Decision Costs In First Price Auctions." *Economic Inquiry* 31:237–244.

Thielmann, I., D.W. Heck, and B.E. Hilbig. 2016. "Anonymity and Incentives: An Investigation of Techniques to Reduce Socially Desirable Responding in the Trust Game." *Judgment and Decision Making* 11:527–536.

Vasilaky, K.N., and J.M. Brock. 2020. "Power(ful) Guidelines for Experimental Economists." *Journal of the Economic Science Association* 6:189–212.

Voslinsky, A., and O.H. Azar. 2021. "Incentives in Experimental Economics." *Journal of Behavioral and Experimental Economics*, pp. 101706.

Wilcox, N.T. 2018. "Conditional Independence in a Binary Choice Experiment." Working paper, Chapman University.

—. 2008. "Stochastic Models for Binary Discrete Choice under Risk: A Critical Primer and Econometric Comparison." In J. C. Cox and G. W. Harrison, eds. *Risk Aversion in Experiments (Research in Experimental Economics, Vol. 12)*. Emerald Group Publishing Limited, Bingley, pp. 197–292.

Woods, D.J. 2021. "Essays on Experimental Economics." PhD thesis, Purdue University.

Yamagishi, T., Y. Li, Y. Matsumoto, and T. Kiyonari. 2016. "Moral Bargain Hunters Purchase Moral Righteousness When it is Cheap: Within-Individual Effect of Stake Size in Economic Games." *Scientific Reports* 6:27824.

Appendices

A Proofs and Derivations

Proof of Proposition 1 in Continuous Case.

I assume that the relevant outcome μ_g is the mean. The required sample size for the t-test for a difference in means is

$$n = 2\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \left(\frac{\sigma}{\mu_T - \mu_C}\right)^2. \tag{A.1}$$

Taking the logarithms, we obtain

$$\ln n = \ln 2 + 2\ln\left(z_{1-\alpha/2} + z_{1-\beta}\right) + 2\ln\sigma - 2\ln\left|\mu_T - \mu_C\right|. \tag{A.2}$$

The derivative of the logarithm of n with respect to τ is

$$(\ln n)' = -2\left[\ln\left|\mu_T - \mu_C\right|\right]'. \tag{A.3}$$

I am assuming that σ does not depend on τ. Notice that if the treatment increases the outcome in the treatment group, $\mu_T' > 0$, then $\mu_T - \mu_C \geqslant 0$. On the other hand, if the treatment decreases the outcome in the treatment group, $\mu_T' < 0$, then $\mu_T - \mu_C \leqslant 0$. Therefore, regardless of whether the treatment increases or decreases the outcome in the treatment group, we have

$$(\ln n)' = -2\frac{\mu_T'}{\mu_T - \mu_C} < 0. \tag{A.4}$$

Increasing the incentives makes the treatment effect larger, which reduces the required sample size.

I assume that the expected per-subject payoff in group g is multiplicative in the treatment and outcome: $\pi_g = \tau_g\mu_g$. The expected per-subject payoff in both groups is

$$\pi_C + \pi_T = \tau_C\mu_C + \tau_T\mu_T = \tau_C\mu_C + \tau_C\mu_T + \tau\mu_T = \tau_C(\mu_C + \mu_T) + \tau\mu_T, \tag{A.5}$$

where I am assuming an additive treatment strength. The derivative of π_T with respect to τ is

$$\pi_T' = (\tau_T\mu_T)' = \mu_T + \tau_T\mu_T' = \mu_T + \tau_C\mu_T' + \tau\mu_T'. \tag{A.6}$$

The derivative of the logarithm of per-subject payoff with respect to τ is

$$[\ln(2w + \pi_C + \pi_T)]' = \frac{\pi_T'}{2w + \pi_C + \pi_T} \tag{A.7}$$

$$= \frac{\mu_T + \tau_C\mu_T' + \tau\mu_T'}{2w + \tau_C(\mu_C + \mu_T) + \tau\mu_T} \tag{A.8}$$

Having the derivatives in (A.4) and (A.7), we can write down the derivative of the logarithm of the total expected budget[34]

$$(\ln \pi)' = (\ln n)' + [\ln(2w + \pi_C + \pi_T)]' \tag{A.9}$$

$$= -2\frac{\mu_T'}{\mu_T - \mu_C} + \frac{\mu_T + \tau_C\mu_T' + \tau\mu_T'}{2w + \tau_C(\mu_C + \mu_T) + \tau\mu_T}. \tag{A.10}$$

Consider first the limiting behavior of the two components of $(\ln \pi)'$ as $\tau \to 0^+$. First note that $\lim_{\tau\to 0^+} \mu_T = \mu_C$. Also, if μ_T is bounded then $\lim_{\tau\to 0^+} \mu_T' < \infty$ by definition, and if μ_T is unbounded, then this is true by assumption. Then the limit of the first component, $(\ln n)'$, is

$$\lim_{\tau\to 0^+} (\ln n)' = -\frac{2\lim_{\tau\to 0^+} \mu_T'}{\lim_{\tau\to 0^+} \mu_T - \mu_C} = -\infty, \tag{A.11}$$

which is true both when μ_T' is positive or negative.

The limit of the second component, $[\ln(2w + \pi_C + \pi_T)]'$, is

$$\lim_{\tau\to 0^+} [\ln(2w + \pi_C + \pi_T)]' = \frac{\mu_C + \tau_C \lim_{\tau\to 0^+} \mu_T'}{2w + 2\tau_C\mu_C}, \tag{A.12}$$

which is finite. Combining the two results, we conclude that

$$\lim_{\tau\to 0^+} (\ln \pi)' = -\infty. \tag{A.13}$$

By continuity of μ_T', we can find a point $a > 0$ arbitrarily close to 0, such that the value of $(\ln \pi)'$ at a is also negative.

Now consider what happens as $\tau \to \infty$. It is useful to split further analysis by whether μ_T' is positive or negative and whether μ_T is bounded or not. Suppose first that $\mu_T' > 0$ and $\lim_{\tau\to\infty} \mu_T < +\infty$. This implies that $\lim_{\tau\to\infty} \tau(\ln \mu_T)' = \lim_{\tau\to\infty} d\ln\mu_T/d\ln\tau = 0$. Consider the following limit:

$$\lim_{\tau\to\infty} \frac{[\ln(2w + \pi_C + \pi_T)]'}{\mu_T'} = \lim_{\tau\to\infty} \frac{\frac{\mu_T}{\tau\mu_T'} + 1 + \frac{\tau_C}{\tau}}{\frac{2w}{\tau} + \frac{\tau_C}{\tau}(\mu_C + \mu_T) + \mu_T} \tag{A.14}$$

$$= \frac{\frac{1}{\lim_{\tau\to\infty} \tau(\ln\mu_T)'} + 1}{\lim_{\tau\to\infty} \mu_T} = +\infty \tag{A.15}$$

Plugging this result into the limit for $(\ln \pi)'$, we get

$$\lim_{\tau\to\infty} (\ln \pi)' = \left(\lim_{\tau\to\infty} \mu_T'\right)\left[-\frac{2}{\lim_{\tau\to\infty} \mu_T - \mu_C} + \lim_{\tau\to\infty} \frac{[\ln(2w + \pi_C + \pi_T)]'}{\mu_T'}\right] \tag{A.16}$$

The term in brackets goes to $+\infty$. By continuity of μ_T', we can find a point b, such that $a < b < \infty$ and the value of the term in brackets is positive. The value of μ_T' at that point is also positive. Therefore, the value of $(\ln \pi)'$ at b is positive.

Now suppose that μ_T is unbounded. Recall that in this case we assume that $\lim_{\tau\to\infty} \tau(\ln\mu_T)' =$

[34] It is easier to work with the logarithm of π than with π itself. The signs of the derivatives of $\ln \pi$ and π with respect to τ, as well as extremum points, are the same.

$\lim_{\tau\to\infty} d\ln\mu_T/d\ln\tau < 1$. Consider the following limit:

$$\lim_{\tau\to\infty}\frac{[\ln(2w+\pi_C+\pi_T)]'}{(\ln\mu_T)'} = \lim_{\tau\to\infty}\frac{\frac{\mu_T}{\tau\mu_T'}+1+\frac{\tau_C}{\tau}}{\frac{2w}{\tau\mu_T}+\frac{\tau_C}{\tau}(\frac{\mu_C}{\mu_T}+1)+1} \tag{A.17}$$

$$= \frac{1}{\lim_{\tau\to\infty}\tau(\ln\mu_T)'}+1 \tag{A.18}$$

Plugging this result into the limit for $(\ln\pi)'$, we get

$$\lim_{\tau\to\infty}(\ln\pi)' = \lim_{\tau\to\infty}(\ln\mu_T)'\left[-\frac{2}{1-\frac{\mu_C}{\mu_T}}+\frac{[\ln(2w+\pi_C+\pi_T)]'}{(\ln\mu_T)'}\right] \tag{A.19}$$

$$= \left(\lim_{\tau\to\infty}(\ln\mu_T)'\right)\left[\frac{1}{\lim_{\tau\to\infty}\tau(\ln\mu_T)'}-1\right] \tag{A.20}$$

The term in brackets is positive since $\lim_{\tau\to\infty}\tau(\ln\mu_T)' < 1$ by assumption. By continuity of μ_T', we can find a point b, such that $a < b < \infty$ and the value of $(\ln\mu_T)'$ at that point is also positive. Therefore, the value of $(\ln\pi)'$ at b is positive.

Now suppose $\mu_T' < 0$. I only consider the case when $\mu_T \geqslant 0$. It is hard to imagine a relevant economic choice variable that would be negative, let alone go to $-\infty$ in the limit. Note that $\tau(\ln\mu_T)' = d\ln\mu_T/d\ln\tau$ is negative and converges to 0 in the limit. Retracing the previous steps, we can conclude that the limit in (A.14) is $-\infty$. What would happen to the limit in (A.16)? The term in the brackets would go to $-\infty$. The term is parenthesis is negative but converges to 0. By continuity of μ_T', we can find a point b, such that $a < b < \infty$ and the value of the term in brackets is negative. The value of μ_T' at that point is also negative. Therefore, the value of $(\ln\pi)'$ at b is positive.

Now we can invoke the Intermediate Value Theorem. Since the value of $(\ln\pi)'$ is negative at a and positive at b, $a < b$, and the function is continuous (since $\mu_T \in C^1$), there must be a point τ^* at which $(\ln\pi)'$ is exactly zero. Moreover, the First Order Sufficient Condition for a Minimum requires the smallest such τ^* to be the minimum point.

QED.

Proof of Proposition 1 in Discrete Case.

I assume that the relevant outcome μ_g is the proportion of subjects choosing a given alternative, labeled A. The required sample size for the test for a difference in proportions is

$$n = (z_{1-\alpha/2}+z_{1-\beta})^2\frac{\mu_T(1-\mu_T)+\mu_C(1-\mu_C)}{(\mu_T-\mu_C)^2}. \tag{A.21}$$

To simplify further notation, I denote $\sigma^2 \equiv \mu_T(1-\mu_T)+\mu_C(1-\mu_C)$. Taking the logarithms on both sides, we obtain

$$\ln n = 2\ln(z_{1-\alpha/2}+z_{1-\beta})+\ln\sigma^2-2\ln|\mu_T-\mu_C|. \tag{A.22}$$

The derivative of the logarithm of n with respect to τ is

$$(\ln n)' = \mu_T'\frac{1-2\mu_T}{\sigma^2}-2\frac{\mu_T'}{\mu_T-\mu_C} \tag{A.23}$$

$$= \mu_T'\left[\frac{1-2\mu_T}{\sigma^2}-\frac{2}{\mu_T-\mu_C}\right]. \tag{A.24}$$

This expression is valid both for the case when $\mu_T' < 0, \mu_T < \mu_C$ and when $\mu_T' > 0, \mu_T > \mu_C$. If $\mu_T < 1/2$ (in case $\mu_T' < 0$) or $\mu_T > 1/2$ (in case $\mu_T' > 0$) an increase in τ leads to a decrease in n.

The expected per-subject payoff in the control group is $\pi_C = \mu_C V_A + (1 - \mu_C)V_B$, where V_A is the expected value of alternative A and V_B is the expected value of all other alternatives. The incentives in the treatment group can be introduced in different ways. One possibility is that only the value of alternative A is scaled by $\tau \geqslant 1$ (multiplicative treatment strength). Another possibility is that the values of all alternatives are scaled, as in Holt and Laury (2002).

I begin by considering the former case first. In this case, the expected per-subject payoff in the treatment group is $\pi_T = \tau\mu_T V_A + (1 - \mu_T)V_B$. The expected per-subject payoff in both groups is $\pi_C + \pi_T = V_A(\mu_C + \tau\mu_T) + V_B(2 - \mu_C - \mu_T)$. The derivative of π_T with respect to τ is $\pi_T' = \mu_T V_A + \mu_T'(\tau V_A - V_B)$. The derivative of the logarithm of the total expected per-subject payoff with respect to τ is

$$[\ln(2w + \pi_C + \pi_T)]' = \frac{\pi_T'}{2w + \pi_C + \pi_T} \tag{A.25}$$

$$= \frac{\mu_T V_A + \mu_T'(\tau V_A - V_B)}{2w + V_A(\mu_C + \tau\mu_T) + V_B(2 - \mu_C - \mu_T)}. \tag{A.26}$$

The derivative of the logarithm of the total expected budget is

$$(\ln \pi)' = (\ln n)' + [\ln(2w + \pi_C + \pi_T)]' \tag{A.27}$$

$$= \mu_T' \left[\frac{1 - 2\mu_T}{\sigma^2} - \frac{2}{\mu_T - \mu_C}\right] + \frac{\mu_T V_A + \mu_T'(\tau V_A - V_B)}{2w + V_A(\mu_C + \tau\mu_T) + V_B(2 - \mu_C - \mu_T)}. \tag{A.28}$$

Consider first the limiting behavior of the two components of $(\ln \pi)'$ as $\tau \to 1^+$. First note that $\lim_{\tau\to 1^+} \mu_T = \mu_C$. Also note that $\lim_{\tau\to 1^+} |\mu_T'| < \infty$ since $\mu_T \in [0, 1]$. Then the limit of the first component, $(\ln n)'$, is

$$\lim_{\tau\to 1^+} (\ln n)' = \left(\lim_{\tau\to 1^+} \mu_T'\right) \left[\frac{1 - 2\mu_C}{2\mu_C(1 - \mu_C)} - \frac{2}{\lim_{\tau\to 1^+}(\mu_T - \mu_C)}\right]. \tag{A.29}$$

In case $\mu_T' > 0, \mu_T > \mu_C$, the limit in parentheses is finite and positive, while the second term in brackets goes to $-\infty$. In case $\mu_T' < 0, \mu_T < \mu_C$, the limit in parentheses is finite and negative, while the second term in brackets goes to $+\infty$. Therefore, $\lim_{\tau\to 1^+}(\ln n)' = -\infty$.

The limit of the second component, $[\ln(2w + \pi_C + \pi_T)]'$, is

$$\lim_{\tau\to 1^+} [\ln(2w + \pi_C + \pi_T)]' = \frac{\mu_C V_A + \lim_{\tau\to 1^+}(\mu_T)'(V_A - V_B)}{2w + 2V_A\mu_C + 2V_B(1 - \mu_C)}, \tag{A.30}$$

which is finite. Combining the two results, we conclude that

$$\lim_{\tau\to 1^+} (\ln \pi)' = -\infty. \tag{A.31}$$

By continuity of μ_T', we can find a point $a > 1$ arbitrarily close to 1, such that the value of $(\ln \pi)'$ is also negative.

Now consider what happens as $\tau \to \infty$. First note that $\lim_{\tau\to\infty} \mu_T$ is either 1 (in case $\mu_T' > 0$) or 0 (in case $\mu_T' < 0$). Also note that $\lim_{\tau\to\infty} \mu_T' = 0$ and $\lim_{\tau\to\infty} \tau(\ln \mu_T)' = \lim_{\tau\to\infty} d\ln\mu_T/d\ln\tau =$

0 since $\mu_T \in [0,1]$. Consider the following limit:

$$\lim_{\tau\to\infty} \frac{[\ln(2w+\pi_C+\pi_T)]'}{\mu_T'} = \lim_{\tau\to\infty} \frac{\frac{\mu_T}{\mu_T'}V_A + \tau V_A - V_B}{2w + V_A(\mu_C+\tau\mu_T) + V_B(2-\mu_C-\mu_T)} \tag{A.32}$$

$$= \lim_{\tau\to\infty} \frac{\frac{\mu_T}{\tau\mu_T'}V_A + V_A - \frac{V_B}{\tau}}{\frac{2w}{\tau} + V_A(\frac{\mu_C}{\tau}+\mu_T) + \frac{V_B}{\tau}(2-\mu_C-\mu_T)} \tag{A.33}$$

$$= \lim_{\tau\to\infty} \frac{\frac{1}{\tau(\ln\mu_T)'}V_A + V_A - \frac{V_B}{\tau}}{\frac{2w}{\tau} + V_A(\frac{\mu_C}{\tau}+\mu_T) + \frac{V_B}{\tau}(2-\mu_C-\mu_T)}. \tag{A.34}$$

Suppose first that $\mu_T' > 0, \mu_T > \mu_C$. Then the limit becomes

$$\lim_{\tau\to\infty} \frac{[\ln(2w+\pi_C+\pi_T)]'}{\mu_T'} = \lim_{\tau\to\infty} \frac{\frac{1}{\tau(\ln\mu_T)'}V_A + V_A}{V_A} \tag{A.35}$$

$$= \frac{1}{\lim_{\tau\to\infty}\tau(\ln\mu_T)'} + 1 \tag{A.36}$$

$$= +\infty. \tag{A.37}$$

In case $\mu_T' < 0, \mu_T < \mu_C$,

$$\lim_{\tau\to\infty} \frac{[\ln(2w+\pi_C+\pi_T)]'}{\mu_T'} = -\infty, \tag{A.38}$$

since the denominator goes to 0 and the numerator goes to $-\infty$.

Then from (A.27), we get

$$\lim_{\tau\to\infty} (\ln\pi)' = \lim_{\tau\to\infty} \left(\mu_T' \left[\frac{1-2\mu_T}{\sigma^2} - \frac{2}{\mu_T-\mu_C} + \frac{[\ln(2w+\pi_C+\pi_T)]'}{\mu_T'}\right]\right). \tag{A.39}$$

Suppose first that $\mu_T' > 0, \mu_T > \mu_C$. Then the limit becomes

$$\lim_{\tau\to\infty} (\ln\pi)' = \lim_{\tau\to\infty} (\mu_T') \left[-\frac{1}{\mu_C(1-\mu_C)} - \frac{2}{1-\mu_C} + \lim_{\tau\to\infty} \frac{[\ln(2w+\pi_C+\pi_T)]'}{\mu_T'}\right]. \tag{A.40}$$

The term in brackets goes to $+\infty$. By continuity of μ_T', we can find a point b, such that $a < b < \infty$ and the value of the term in brackets is positive. The value of μ_T' at that point is also positive. Therefore, the value of $(\ln\pi)'$ at b is positive.

On the other hand, if $\mu_T' < 0, \mu_T < \mu_C$, the limit becomes

$$\lim_{\tau\to\infty} (\ln\pi)' = \lim_{\tau\to\infty} (\mu_T') \left[\frac{1}{\mu_C(1-\mu_C)} + \frac{2}{\mu_C} + \lim_{\tau\to\infty} \frac{[\ln(2w+\pi_C+\pi_T)]'}{\mu_T'}\right]. \tag{A.41}$$

The term in brackets goes to $-\infty$. Following the same logic, in this case we can find a point b, such that $a < b < \infty$ and the value of the term in brackets is negative. The value of μ_T' at that point is also negative. Therefore, the value of $(\ln\pi)'$ at b is positive. In both cases, therefore, we have that the value of $(\ln\pi)'$ at b is positive.

Now we can invoke the Intermediate Value Theorem. Since the value of $(\ln\pi)'$ is negative at a and positive at b, $a < b$, and the function is continuous (since $\mu_T \in C^1$), there must be a point τ^* at which $(\ln\pi)'$ is exactly zero. Moreover, the First Order Sufficient Condition for a Minimum requires the smallest such τ^* to be the minimum point.

Another possibility for introducing incentives is to scale all alternatives by τ. In that case, $\pi_T = \tau(\mu_T V_A + (1 - \mu_T V_B))$. This will lead to the following formula for the derivative of the logarithm of the total expected per-subject payoff with respect to τ:

$$[\ln(2w + \pi_C + \pi_T)]' = \frac{\pi_T'}{2w + \pi_C + \pi_T} \tag{A.42}$$

$$= \frac{V_B + \Delta V(\tau\mu_T' + \mu_T)}{2w + V_B 1 + \tau) + \Delta V(\mu_C + \tau\mu_T)}, \tag{A.43}$$

where $\Delta V \equiv V_A - V_B$. The limit of μ_T as $\tau \to 1^+$ will still be equal to μ_C. However, the limit as $\tau \to \infty$ will be 1/[number of alternatives], since scaling every alternative eventually will make them equally attractive. Applying the same steps as we did previously, it is easy to show that the limits of $(\ln \pi)'$ at each endpoint will still be of the opposite signs, and the result will hold.

QED.