Measuring and Bounding Experimenter Demand*

Jonathan de Quidt[†] Johannes Haushofer[‡]
Christopher Roth[§]

Preliminary and incomplete.

This version: March 20, 2017

Abstract

We propose a technique for assessing robustness of behavioral measures and treatment effects to experimenter demand effects. The premise is that by deliberately inducing demand in a structured way we can measure its influence and construct bounds on demand-free behavior. We provide formal restrictions on choice that validate our method, and a Bayesian model that microfounds them. Six pre-registered experiments with eleven canonical laboratory games and around 18,000 participants demonstrate the technique. We also illustrate how demand sensitivity varies by task, subject pool, gender, real versus hypothetical incentives, and subject attentiveness, and provide both reduced-form and structural analyses of demand effects.

Keywords: Experimenter Demand, Beliefs, Bounding
JEL Classification: B41, C91, C92

*We are grateful to Johannes Abeler, Stefano Caria, Rachel Cassidy, Tom Cunningham, Elwyn Davies, Stefano DellaVigna, Thomas Graeber, Don Green, Alexis Grigorieff, Johannes Hermle, Simon Quinn, Matthew Rabin, Gautam Rao, Bertil Tungodden and Liad Weiss for comments. We thank Stefano DellaVigna, Lukas Kiessling, and Devin Pope for sharing code. Moreover, we would like to thank seminar participants at Bergen, Berlin, Busara, Oxford, Stockholm, Wharton and Wisconsin. We thank Justin Abraham for excellent research assistance.

[†]Institute for International Economic Studies (IIES); e-mail: jonathan.dequidt@iies.su.se.

[‡]Princeton University and Busara Center for Behavioral Economics; e-mail: haushofer@princeton.edu.

[§]Department of Economics, University of Oxford and CSAE; e-mail: christopher.roth@economics.ox.ac.uk.

1 Introduction

A basic concern in experimental work with human subjects is that, knowing that she is being experimented on, the subject may change her behavior. Specifically, subjects may try to infer the experimenter's objective from their treatment, and then act accordingly (Rosenthal, 1966; Zizzo, 2010). For instance, subjects who believe the experimenter wants to show that people free-ride in public good games might play more selfishly than they otherwise would. Thus, instead of measuring the subject's "natural" choice, the experimental data are biased by an unobservable *experimenter demand effect*. Demand effects pose a threat to external validity, because participants would make different choices if the experimenter were absent, and shroud the interpretation of treatment effects.

The core idea of our paper is that it is possible to bound demand-free behavior and treatment effects by deliberately *inducing* experimenter demand and measuring its influence. For example, in an effort task we might tell some participants "you will do us a favor if you work more than you normally would," and others "you will do us a favor if you work less than you normally would." Under the identifying assumption that any underlying demand effect is less extreme (in a sense that we will formalize) than our manipulations, choices under these instructions give upper and lower bounds on demand-free behavior, and by combining bounds from different experimental treatments we can estimate bounds on treatment effects.

We begin with a simple Bayesian model of decision-making that motivates our approach. In our model an experiment defines a mapping from actions to payoffs, and subjects choose a single action to maximize utility. The experimenter is only interested in measuring the "natural" action that maximizes the subject's utility as derived from the experimental payoffs.

However, the subject is also motivated to take actions he perceives will "please" the experimenter by conforming to her research objectives. He tries to infer those objectives from the design features and distorts his action accordingly, biasing the experimenter's estimates. Our demand treatments attempt to manipulate those beliefs in opposing directions, so as to identify an interval that contains the natural action. We remain agnostic about *why* the subject wishes to please the experimenter; possible motives include altruism, a motive to conform, or a belief that he will ultimately be rewarded for doing so.

We provide an extensive set of applications of the method. We conduct six online experiments with approximately 18,000 participants in total, in which we construct bounds on demand-free behavior for 11 canonical games and preference measures.[1] In each game we employ positive and negative "demand treatments" which tell participants that they will "do us a favor" if they choose a higher or lower action than they normally would.

Responses to our demand treatments are significant and variable across tasks. The difference in average (standardized) behavior between our positive and negative demand treatment groups ranges from approximately 0.25 standard deviations for incentivized real effort to around 1 standard deviation for trust game second movers. In an application to treatment effect estimation, we also derive bounds on the real effort response to performance pay. The bounds we obtain exclude zero, but are quite wide, ranging from a 0.25 to a 1.35 standard deviation increase in effort.

A concern in applying our method is that participants might defy the experimenter's wishes, doing the opposite of what is demanded. Unlike the estimation of local average treatment effects (Angrist and Imbens, 1994), we

[1]Specifically, we study simple time, risk and ambiguity preference elicitation tasks, a real effort task with and without performance incentives, a lying game, dictator game, ultimatum game (first and second mover) and trust game (first and second mover). Our data come from a US nationally representative online panel as well as US-based Amazon Mechanical Turk (MTurk) participants.

do not require a strict "no defiers" assumption; we can still obtain bounds on average actions and treatment effects provided there are not too many defiers. In support of this hypothesis, we show experimentally that our manipulations do increase and decrease average actions as intended, relative to treatment groups not exposed to any demand manipulation.[2]We interpret these findings as evidence that the method functions in line with the theory, and that while we cannot rule out some defier behavior, it is not sufficient to overwhelm the compliers. We do emphasize however that this is not a test of our core identifying assumption, which depends on unobservable demand-free choices.

Next, following the basic approach of DellaVigna and Pope (2016b) we illustrate how our demand treatments can be used in conjunction with a structural model to a) estimate a monetary value for the representative subject's utility of conforming to the experimenter's wishes, and b) obtain unconfounded estimates of the main structural parameters of interest.

One implication of the theory is a trade-off in applying our method. Strong demand manipulations provide reliable but wide bounds on demand-free behavior, because they shift participants' beliefs a lot even if the underlying influence of demand is small. Weaker manipulations have the potential to tighten bounds, at the cost of decreased robustness. In another series of experiments on MTurk, we employ demand treatments in which we signal an experimental hypothesis to subjects, without explicitly demanding that they conform to it. Specifically, we tell them "we expect that participants who are shown these instructions will [work, invest, . . .] more/less than they normally would." We find a much more moderate response to these treatments, and consequently obtain much narrower bounds, ranging from around 0.1 to 0.3 standard deviations.

[2]We measure actions without demand manipulations for a subset of tasks: time, risk, effort and the dictator game.

Finally, we examine several moderators of sensitivity to experimenter demand. First, we find that women generally respond more to our strong demand treatments than men. Second, surprisingly, we find that varying whether choices are incentivized or hypothetical does not affect sensitivity to our demand manipulations. Third, we find some evidence that more attentive respondents, as measured by a screener task at the beginning of the experiment, respond more strongly. Fourth, we compare behavior between our two subject pools—Amazon MTurk workers and a US representative online panel—and find that both are equally responsive.

We contribute to the small literature discussing experimenter demand effects (Shmaya and Yariv, 2016; Zizzo, 2010; Zizzo and Fleming, 2011), demand characteristics (Orne, 1962), and obedience to the experimenter (Milgram, 1963). List (2007) and Bardsley (2008) show that adding the option to take money in a dictator game dramatically reduces giving. They argue that behavior in the dictator game is to a large degree an artefact of the experimental situation.

We also contribute to the literature which examines the effects of anonymity on social behavior in the laboratory (Barmettler et al., 2012; Levitt and List, 2007; Hoffman et al., 1996, 1994). Barmettler et al. (2012) find no evidence that experimenter-subject anonymity affects behavior in standard social preference measures, while other studies document that non-anonymity in the lab can increase pro-social behavior (List et al., 2004; Masclet et al., 2003). We also relate to work that explores the principal-agent relationship between experimenter and subject (Chassang et al., 2012; Shmaya and Yariv, 2016).

Second, our paper contributes to the literature discussing whether lab behavior generalizes to the field (Levitt and List, 2007; Harrison and List, 2004). List (2006) finds that behavior in the lab environment can be at odds with behavior in the field, which could be due to differences in demand effects in the lab as opposed to the field.

Third, we relate to the growing literature on the effects of social pressure on economic, and social behavior, for example, charitable giving (DellaVigna et al., 2012) and voting (DellaVigna et al., 2017; Gerber et al., 2008).[3]

The paper proceeds as follows: in section 2 we set up a simple theoretical model of experimenter demand that motivates our approach. In section 3 we describe the data and our experimental design. In section 4, we present experimental results and structural estimates. Section 5 concludes. A set of web appendices provide additional results and theory.

2 Theory

We model a decision-maker (he) who has preferences over the outcomes induced by his action $a \in \mathbb{R}$ in an experiment. a can be continuous or discrete, but for simplicity we focus on the case of continuous actions with a natural ordering (more/less effort, investment, giving). The analysis extends naturally to the case where a is the probability of a binary choice. While throughout the analysis we treat a as the choice of a representative agent, it is straightforward to reinterpret as a population or group mean action, and our conditions as applying to average actions.

In the absence of an experimenter, the optimal action is simply a function of the decision-making *environment*. We index environments by $\zeta \in Z$, where ζ captures aspects including subject characteristics (e.g. male/female, student/representative sample), setting (e.g. lab/field, USA/Kenya, online/in-person), experimental treatments, the units in which a is measured, and so on. Given ζ, the optimal "natural" (experimenter-absent) action is $a(\zeta)$.

[3]This literature is also linked to work on moral suasion and pro-social behavior (Dal Bó and Dal Bó, 2014).

The experimenter (she) is interested in either i) measuring a specific action $a(\zeta)$ (e.g., the level of giving out of an endowment), or ii) a treatment effect $a(\zeta_1) - a(\zeta_0)$ (e.g., the effect of incentives on effort provision). Unfortunately, her task is complicated by experimenter demand. Knowing that he is a participant in an experiment, the decision-maker changes his action according to his belief about the experimenter's wishes or objectives. Instead of $a(\zeta)$, he chooses action $a^L(\zeta)$ where L signifies the presence of a "latent", unobserved experimenter demand influence. The influence could increase or decrease a: $a^L(\zeta) \gtreqless a(\zeta)$. We define the *latent demand effect* in environment ζ as the difference $a^L(\zeta) - a(\zeta)$.

While nonzero latent demand automatically biases estimates of mean actions, it does not necessarily bias estimates of treatment effects. The logic of a randomized experiment is to induce orthogonal variation in a treatment so as to estimate its influence purged of confounds. If the influence of latent demand is orthogonal to the treatment, the treatment effect is not biased. To see this, note that the treatment effect can be decomposed as follows:

$$a^L(\zeta_1) - a^L(\zeta_0) = \underbrace{a(\zeta_1) - a(\zeta_0)}_{\text{Effect of interest}} + \underbrace{[a^L(\zeta_1) - a(\zeta_1)]}_{\text{Latent demand in } \zeta_1} - \underbrace{[a^L(\zeta_0) - a(\zeta_0)]}_{\text{Latent demand in } \zeta_0} \quad (1)$$

The first term on the right-hand side is the treatment effect of interest. The second and third capture the potential bias due to experimenter demand. If both demand effects are equal they cancel and the treatment effect is identified, but they may not cancel, either because the subject's inference or his response to a given inference (or both) varies with ζ.

Example 1. Consider two variants on the classic Dictator game, in which the subject is told to choose what fraction of \$10 to give to another participant. In variant 0, she is told that the recipient is aware that the choice is taking place, while in variant 1 they are unaware (for instance, the money will just be added to a show-up fee). In both scenarios, absent any motive for pleasing the experimenter she would prefer to give \$4, so the true treatment effect is $a(\zeta_1) - a(\zeta_0) = \$0$. However, in variant 0 she infers that the experimenter wants her to be generous, so she gives \$5, while in variant 1 she infers that the experimenter wants her to be selfish, so she gives zero. Then $a^L(\zeta_0) - a(\zeta_0) = \$1$ and $a^L(\zeta_1) - a(\zeta_1) = -\$4$, so $a^L(\zeta_1) - a^L(\zeta_0) = -\$5$ and we spuriously identify a treatment effect that is in reality a demand effect.

2.1 Demand treatments

We now assume that the experimenter has at her disposal a particular kind of treatment manipulation, $d^T \in \{-1, 1, \emptyset\}$ which we call a *demand treatment*. $d^T = -1$ is a manipulation that deliberately signals a demand that the decision-maker decrease his action, $d^T = 1$ demands that he increase. $d^T = \emptyset$ if the experimenter does not use a demand treatment, as in the vast majority of experimental designs. The actions induced by the demand treatments are $a^-(\zeta)$ and $a^+(\zeta)$ (and $a^L(\zeta)$ if $d^T = \emptyset$). For illustrative purposes we assume that there exists just one type of positive and negative demand treatment, and discuss treatments that differ in intensity below.

Our first substantive assumption is a basic monotonicity condition on the response to d^T:

Assumption 1. *Monotone demand treatment effects* $a^-(z) \leq a^L(z) \leq a^+(z), \forall z \in Z$.

Assumption 1 requires that a deliberate attempt by the experimenter to demand an increase in the action does not decrease it, and vice versa. It has a natural counterpart in the literature on local average treatment effects (Angrist and Imbens, 1994): the assumption rules out "defier" behavior whereby participants demanded to increase their actions, decrease them, and vice versa. While this is a strong assumption, we note that we only require it to hold for *average* actions, which is weaker than the standard no-defiers assumption. Moreover, Assumption 1 is testable (for average behavior) at $t = \zeta$ because we can test whether a^-, a^L and a^+ are correctly ordered. We perform this test for some of our applications, discussed below.

Our next assumption is central to our bounding exercises, and simply amounts to assuming that the demand treatments are capable of bounding the natural action of interest:

Assumption 2. *Bounding* $a^-(\zeta) \leq a(\zeta) \leq a^+(\zeta)$.

Assumption 2 allows us to bound estimates of mean actions and treatment effects. It implies the following:

$$a(\zeta) \in [a^-(\zeta), a^+(\zeta)] \tag{2}$$

$$a(\zeta_1) - a(\zeta_0) \in [a^-(\zeta_1) - a^+(\zeta_0), a^+(\zeta_1) - a^-(\zeta_0)] \tag{3}$$

This assumption therefore delivers the central result than we can use demand-inducing treatments to obtain bounds on mean actions (equation 2) and treatment effects (equation 3).

These bounds are the main objects of interest for our analysis, but for some purposes we may wish to be able to make comparative statements about demand in different environments. Although the latent demand effect is unobservable, the sensitivity of behavior to d^T may be informative about it. Now we provide an assumption that enables us to make such statements.

First, we define "sensitivity" to be the difference in actions under positive and negative demand treatments: $S(\zeta) = a^+(\zeta) - a^-(\zeta)$.

Remark 1. In addition to bounding the natural action, assumptions 1 and 2 jointly imply that sensitivity $S(\tau)$ provides an upper bound on the magnitude of the latent demand effect: $S(\zeta) \geq \left|a^L(\zeta) - a(\zeta)\right|$.

Proof. Assumption 1 gives $a^L(\zeta) \in [a^-(\zeta), a^+(\zeta)]$ while assumption 2 gives $a(\zeta) \in [a^-(\zeta), a^+(\zeta)]$. Taken together this implies $\left|a^L(\zeta) - a(\zeta)\right| \leq a^+(\zeta) - a^-(\zeta)$. □

This fact enables us to use sensitivity $S(\zeta)$ to make statements of comparative ignorance, in the sense that if $S(\zeta_1) > S(\zeta_0)$ there is more scope for large latent demand effects under ζ_1 than under ζ_0 (or equivalently, our bounds on $a(\zeta_1)$ are wider than those on $a(\zeta_0)$).

However, as sensitivity only gives us an upper bound on the latent demand effect, it could be that the true latent demand effect is larger under ζ_0. Our third assumption, Monotone Sensitivity, allows us to make comparative statements about magnitudes.

Definition. Comparison classes. A comparison class $Z^C \subseteq Z$ is a set of environments for which Monotone Sensitivity holds for all $z \in Z^C$.

Assumption 3. *Monotone Sensitivity $S(z)$ is strictly increasing in $\left|a^L(z) - a(z)\right|$ for all $z \in Z^C$.*

This assumption allows us to make comparative statements about latent demand effects between environments. Natural candidates for comparison classes include environments that differ only in a small number of attributes. For instance, an experimenter might be interested in test whether demand effects are larger in one subject pool compared to another, or under one incentive scheme compared to another.[4] We derive some comparison classes below in the more structured context of our Bayesian model.[5]

[4]A necessary condition for comparisons to be interesting is that actions are measured in the same units. Translating actions into a common scale (for example, "standardizing" the data to measure actions as multiples of the standard deviation) is one way to achieve this. An environment can belong to multiple comparison classes, and all comparison classes may be singletons.

[5]Some experimenters might be willing to assume monotone sensitivity without bounding. In that case the demand treatments can be informative about the scale of experimenter demand in environment ζ without bounding the true action (e.g. it could be that $S(\zeta)$ is some fixed proportion of $\left|a^L(\zeta) - a(\zeta)\right|$). Then, a natural use of

Finally, we describe how demand treatments might be used to extract estimates of a true treatment effect of interest. If the researcher is willing to assume i) monotone sensitivity, and ii) that the latent demand effects under ζ_1 and ζ_0 have the same sign (i.e, $a^L(\zeta_1) - a(\zeta_1) \geq 0 \Leftrightarrow a^L(\zeta_0) - a(\zeta_0) \geq 0$), then, if the following difference-in-differences condition holds: $(a^+(\zeta_1) - a^-(\zeta_1)) - (a^+(\zeta_0) - a^-(\zeta_0)) = 0$, $a^L(\zeta_1) - a^L(\zeta_0)$ identifies the true treatment effect. Even if the condition does not hold, it can be used to sign the bias due to demand. Suppose the condition is positive, implying $S(\zeta_1) > S(\zeta_0)$. Monotone sensitivity then implies that $\left|a^L(\zeta_1) - a(\zeta_1)\right| > \left|a^L(\zeta_0) - a(\zeta_0)\right|$. Then, knowing the sign of the latent demand effect under ζ_1 enables the researcher to sign the bias due to experimenter demand: If $a^L(\zeta_1) - a(\zeta_1) > 0$ the experiment overestimates the treatment effect of interest, if $a^L(\zeta_1) - a(\zeta_1) < 0$ it underestimates it.

2.2 Bayesian model

In this section we provide a simple model of a decision-maker who is subject to demand effects, so as to provide intuition for our main assumptions and precise conditions under which they will or will not hold.

The environment ζ determines the mapping from actions $a \in \mathbb{R}$ into outcomes or distributions over outcomes, over which the decision-maker has preferences. We compactly describe his payoff from action a in environment ζ by $v(a, \zeta)$ where v captures both the payoff structure (mapping from actions to outcomes) and preferences (mapping from outcomes to utility). We assume that v is strictly concave and differentiable, so the natural action $a(\zeta)$ solves $v_1(a(\zeta), \zeta) = 0$.

$S(\zeta)$ is to provide information about "how bad" experimenter demand would have to be to threaten a certain interpretation of the data. She might compute objects like $m = \left|a^L(\zeta) - \alpha\right| / S(\zeta)$ in order to make statements like "latent demand would have to be m multiples of $S(\zeta)$ to be consistent with $a(\zeta) = \alpha$." Analogous approaches exist for bounding bias due to sample selection or violation of the exclusion restriction in IV, eg. Conley et al. (2012), Nevo and Rosen (2012) and Altonji et al. (2005).

Example 2. Effort provision: A risk-neutral decision-maker chooses effort a. Effort is rewarded by a piece-rate ζ and the cost of effort is $c(a)$. Then $v(a, \zeta) = \zeta a - c(a)$ and $a(\zeta) = c'^{-1}(\zeta)$.

2.2.1 Latent demand

Demand enters preferences as follows. Upon observing the experiment and treatment, the decision-maker makes an inference about an unobservable parameter, $h \in \{-1, 1\}$. If $h = -1$, he believes the experimenter benefits from him taking low actions, while if $h = 1$ he believes she benefits from high actions.[6] His preference for pleasing the experimenter is captured by a preference parameter ϕ, which we allow to depend upon ζ, having in mind that ϕ might depend on the identity of the experimenter (e.g. the decision-maker might have different attitudes toward a researcher and a firm) or decision-maker (e.g. women might have different attitudes than men). $\phi(\zeta)$ might also vary with other environment features such as the salience of the potential benefit to the experimenter or how important the subject believes his actions are in achieving the experimenter's objectives. We remain agnostic about *why* the subject wishes to please the experimenter; possible motives include altruism, a perhaps misguided desire to contribute to science, a motive to conform, or a belief that he will ultimately be rewarded for doing so. See e.g. Orne (1962) for discussion.

We assume utility is separable in payoffs and demand, in the following form:

$$U(a, \zeta) = v(a, \zeta) + a\phi(\zeta)E[h|\zeta] \tag{4}$$

[6]We think of the decision-maker perceiving the experimenter's preference over his actions directly, preferring actions toward one or other extreme, rather than the experimental outcomes outcomes induced by those actions. This is equivalent to the case where the outcome is a monotone function of the action and the experimenter's payoff is increasing or decreasing in the experimental outcome.

where $E[h|\zeta] = Pr(h = 1|\zeta) \times 1 + Pr(h = -1|\zeta) \times (-1) = 2Pr(h = 1|\zeta) - 1$. The optimal action $a^L(\zeta)$ solves:

$$v_1(a^L(\zeta), \zeta) + \phi E[h|\zeta] = 0 \tag{5}$$

so $a^L(\zeta) = a(\zeta) \Leftrightarrow \phi E[h|\zeta] = 0$. There is therefore no demand confound if either a) the decision-maker assigns equal likelihood to the preferred action being high or low ($E[h|\zeta] = 0$), or when he does not care about the experimenter's objectives ($\phi = 0$).

We assume that the decision-maker's prior over h is $h_0 = 0$, so in the absence of any new information about h he chooses $a(\zeta)$. The relation between actions and beliefs is captured by $da^L(\zeta)/dE[h|\zeta] = -\phi/v''(a, \zeta)$, which has the same sign as ϕ.

We model the decision-maker's learning about h as follows. The environment ζ includes a signal $h^L(\zeta) \in \{-1, 1\}$ which the decision-maker believes is a sufficient statistic for h, i.e. it contains all of the information in ζ about h, so $E[h|\zeta] = E[h|h^L(\zeta)]$. He believes that with probability $p^L(\zeta)$, the signal is correct ($h^L = h$) and with probability $1 - p^L(\zeta)$, it is pure noise ($h^L = \epsilon$, where ϵ equals negative or positive one with equal probability). We impose that $p^L(\zeta) \in [0, 1)$, so the latent demand signal can never be perfectly informative.[7] It is straightforward to see (and we show in the Appendix) that:

$$E[h|h^L(\zeta)] = h^L(\zeta)p^L(\zeta) \tag{6}$$

The decision-maker's belief depends on the experimental treatment in two ways. First, the treatment determines the sign of $h^L(\zeta)$, i.e. whether it induces a belief that the experimenter wants a high or low action. This

[7] This assumption avoids the possibility that both latent demand and the demand treatment are seen as perfectly informative, but contradictory.

determines the *direction* of the latent demand effect. Second, the *strength* of the latent demand effect depends on $p^L(\zeta)$, which measures the informativeness of the signal to be about h. [8]

2.2.2 Demand treatments

We now assume that the experimenter has the option to send a "demand treatment" signal $h^T \in \{-1, 1, \emptyset\}$, which is either positive ($h^T = 1$), negative ($h^T = -1$), or no signal ($h^T = \emptyset$). These signals deliberately direct the decision-maker toward a high or low action by changing his belief about h. We assume that if the experimenter does not send a signal (no demand treatment), the decision-maker does not update his belief about h, i.e. he does not draw any inference from the absence of a demand treatment. This assumption is reasonable as at present demand treatments are rarely used in experiments and therefore presumably not expected.

The decision-maker believes that h^T is informative about h: with probability p^T, h^T equals h, and with probability $1 - p^T$ it equals η, which takes values negative and positive one with equal probability. η and ϵ are believed to be independent (we return to this assumption below). Based on h^T, the decision-maker updates his beliefs about h. We show in the Appendix that the Bayesian posterior is:

$$E[h|h^T, h^L(\zeta)] = \frac{h^L(\zeta)p^L(\zeta) + h^T p^T}{1 + h^L(\zeta)p^L(\zeta)h^T p^T} \tag{7}$$

Thus, if $h^L(\zeta) = h^T$, the demand treatment reinforces the subject's belief, while if the signals have opposite signs they offset one another.

[8] As an aside, we note that the setup nests "natural field experiments" (Harrison and List, 2004) that induce no demand because the subject does not know they are in an experiment. Such experiments can be viewed as setting $p^L(\zeta) = 0$ (so the treatment is not perceived to be informative about h and therefore $E[h|h^L(\zeta)] = 0$), or $\phi(\zeta) = 0$.

2.2.3 Assumptions

We now provide the formal connection from the Bayesian model to the assumptions outlined in Section 2.1.

First, Assumption 1 (monotone demand treatment effects) states that a positive demand treatment increases the action (relative to no demand treatment) and the negative demand treatment decreases it. In the Bayesian model the conditions for this relationship to hold are $\phi(E[h|h^T = 1, h^L(\zeta)] - E[h|h^L]) \geq 0$ and $\phi(E[h|h^T = -1, h^L(\zeta)] - E[h|h^L]) \leq 0$. It is straightforward to see (and we show in the Appendix) that except for the trivial case $p^T = 0$, these conditions are satisfied if and only if $\phi \geq 0$, i.e. the subject has a weak preference for pleasing the experimenter (and therefore does not "defy" the perceived demand).

Proposition 1. *Monotone demand treatment effects Assumption 1 holds for all p^T if and only if $\phi \geq 0$.*

Second, Assumption 2 (bounding) states that the demand treatments provide bounds on the true action. In the Bayesian model, the action is larger or smaller than $a(\zeta)$ when $\phi E[h|h^T, h^L] \gtreqless 0$. Given that $\phi \geq 0$ by Assumption 1, we need $E[h|h^T = 1, h^L] \geq 0$ and $E[h|h^T = -1, h^L] \leq 0$. This is obviously guaranteed if h^T and h^L have the same sign, so we simply need to check whether it holds when the demand treatment and latent demand are in opposite directions, i.e. $E[h|h^T = 1, h^L = -1] \geq 0$ and $E[h|h^T = -1, h^L = 1] \leq 0$. Given our restriction $p^L(\zeta) < 1$, inspection of (7) reveals that these conditions hold if and only if $p^T \geq p^L(\zeta)$, i.e. the the decision-maker perceives the demand treatment as at least as informative about h as the latent demand signal.

Proposition 2. *Bounding Assumption 2 holds if and only if $p^T \geq p^L(\zeta)$.*

Finally, Assumption 3 (monotone sensitivity) states that within a comparison class Z^C of environments, differences in sensitivity are informative about differences in underlying latent demand. Since latent demand and sensitivity can vary for multiple reasons, there is no simple condition that guarantees when this assumption will and will not hold. In Appendix C.4 we work out the following cases:

1. We show that monotone sensitivity holds when variation in demand effects is driven by differences in the strength of preference for pleasing the experimenter, ϕ.

2. We analyze monotone sensitivity when variation in demand effects is driven by differences in the utility function, v, deriving specific conditions when v is additively or multiplicatively separable and providing examples.

3. We show that monotone sensitivity holds for variation driven by inattention to experimenter demand.

4. We show that monotone sensitivity does *not* hold in general for variation driven by differences in beliefs.

We make use of these findings when interpreting the results of our empirical analyses of heterogeneity.

2.3 Discussion and extensions

2.3.1 "Strong" and "weak" demand treatments

As shown above, the bounding assumption holds when $p^T \geq p^L(\zeta)$. Thus far we have assumed that there is only one demand treatment (with a positive and negative variant), but in reality there are many different ways to signal a desire for high or low actions. How should the experimenter choose?

Observe that the width of the bounds $[a^-(\zeta), a^+(\zeta)]$ is increasing in p^T.[9] Therefore the tightest bounds, subject to satisfying the bounding condition, are obtained when $p^T = p^L(\zeta)$. In other words, we want the "least informative" demand treatment, conditional on it being more informative than the latent signal. Thus, there exists a trade-off between the informativeness of the demand signal and the tightness of bounds: one can use demand treatments that strongly signal the experimenter's objective, giving confidence in the (wide) bounds obtained, or use more subtle manipulations to obtain tighter bounds, at the risk of failing to bound the true action or treatment effect.

In our applications, we use two demand treatments. The first, "strong" treatment explicitly tells participants what we wish: "You will do us a favor if you [...] more/less than you normally would." The second, "weak" treatment hints at a hypothesis, but does not explicitly tell the subject what we want them to do: "We expect that participants who are shown these instructions will [...] more/less than they normally would." We view the first treatment as being more informative about the experimenter's wishes (i.e., carrying a higher p^T) than the second, therefore generating more reliable but wider bounds.

2.3.2 Fewer treatments

It may not always be necessary to construct two-sided bounds. One such case is when the researcher has a strong prior about the direction of latent demand. For example, if they believe $a^L(\zeta) < a(\zeta)$ they might use only a positive demand treatment and construct bounds $[a^L(\zeta), a^+(\zeta)]$. Alternatively, they may only be interested in one bound. For example, if they only wish to obtain a lower bound on a treatment effect $a(\zeta_1) - a(\zeta_0)$, they

[9] Proof: $dE[h|h^T, h^L]/dp^T = h^T \left(1 - h^{L2} p^{L2}\right) \left(1 + h^L p^L h^T p^T\right)^{-2}$, which has the same sign as h^T, so a^+ is increasing and a^- decreasing in p^T.

might measure only $a^+(\zeta_0)$ and $a^-(\zeta_1)$.

2.3.3 Discrete actions

The analysis easily extends to discrete and possibly un-ordered actions, such as accepting/rejecting a contract, selecting a lottery from a choice list or choosing bundles from a menu. Typically the experimenter will then be interested in the probability that a given option is chosen, and may be concerned that this probability is influenced by participants' beliefs about what the experimenter *wants* them to choose. Demand treatments can be readily constructed to manipulate those beliefs and obtain bounds, for instance telling subjects "you will do us a favor if you (do not) choose option j."

2.3.4 Heterogeneity

Thus far we assumed a representative agent and made assumptions about his behavior. However, the approach naturally extends to the case where participants are heterogeneous and the experimenter is interested in average behavior or average treatment effects. If our non-parametric assumptions 1 and 2 hold for all agents individually, then we can simply reinterpret the natural action a and observed actions a^L, a^+ and a^- as representing mean behaviors and our approach remains valid. Intuitively, if we can bound all individuals' natural actions, then we also bound the mean of those actions.

An important source of heterogeneity is in participants' beliefs about the experimenter's wishes, whereby $E_i[h|h^L]$ takes on different values for different individuals i. This could be because $p_i^L(\zeta)$ (perceived precision of the signal) varies across individuals, $h_i^L(\zeta)$ (perceived direction of demand), or both. However, since the bounding condition depends only on $p^T \geq p^L(\zeta)$ and not the direction h^L, provided the inequality is satisfied each individual's natural action will be bounded by a^- and a^+ and the average natural action is bounded. A simple sufficient condition that guarantees bounding is $p^T \geq \max_i p_i^L(\zeta)$.

2.3.5 Defiers

Our monotone demand treatment effects assumption requires that there are no "defiers" who do the opposite of what they believe the experimenter wishes. In the model, such individuals possess a $\phi < 1$. Bounding then fails for these individuals, because $a^- \geq a^+$. However, we may still be able to bound population average behavior if the number of defiers and their responsiveness are relatively small.

To illustrate, we focus on the special case where preferences v and beliefs $E[h|h^L]$, $E[h|h^T, h^L]$ are identical for all individuals. In other words, all participants believe they know what the experimenter wants, but vary in their desire to conform. This means that the natural action is the same for all individuals and given by the solution to $v_1(a(\zeta), \zeta) = 0$. We label the beliefs H^L, H^+ and H^-. We further restrict preferences to be quadratic in actions, so $v_1(a, \zeta) = b - ca$ where b and c are constants that may depend on ζ. Normalizing c to equal 1, the natural action is equal to b for all individuals, while the action when beliefs equal H is $b + \phi H$. We therefore have:

$$
\begin{aligned}
Ea^L &= b + H^L E\phi_i \\
Ea^+ &= b + H^+ E\phi_i \\
Ea^- &= b + H^- E\phi_i
\end{aligned}
$$

Monotone demand treatment effects is satisfied on average if $Ee^L \in [Ea^-, Ea^+]$, which holds if and only if $E\phi_i \geq 0$. Therefore, testing for monotone demand treatment effects is equivalent to testing whether ϕ is positive on average. Bounding will hold provided $H^{+}E\phi_i \geq 0$ and $H^{-}E\phi_i \leq 0$, which follows if $p^T \geq p^L(\zeta)$ and $E\phi_i \geq 0$.

This case is simple because of the quadratic preferences assumption, which implies that compliers and defiers respond symmetrically in opposite directions. More generally we would require conditions on the joint distribution of preferences and beliefs such that that the response by the compliers "outweighs" that of the defiers.

2.3.6 Extension: learning about ϕ

An interpretation of our demand treatments is that they signal not only the direction of the experimenter's objective, but the salience or intensity of her preference over objectives. For instance "do me a favor" suggests that the choice is important. In appendix C.5 we extend the model to incorporate this feature. We assume that the decision-maker responds more strongly to experimenter demand when they believe that complying with the objective it is more important, and that this belief depends both on latent demand and the demand treatments. We show that the key condition for bounding is still $p^T \geq p^L$, but that demand treatments that are perceived to signal more importance generate wider bounds.

2.3.7 Extension: richer beliefs and correlated signals

Researchers sometimes give experimental participants instructions like "there are no right or wrong answers" or "we are only interested in what you think is the best choice." Such instructions can be naturally thought of as a demand treatment that attempts to demand participants choose the natural action, $a(\zeta)$. It is straightforward to analyze such treatments in our framework. In Appendix C.6 we extend the model to allow h to take three values: $\{-1, 0, 1\}$, where $h = 0$ captures the case where the experimenter wants the subject to choose the natural action.

We show that unless the demand treatment is perceived as fully informative ($p^T = 1$), signaling $h^T = 0$ does *not* induce the subject to take the natural action, i.e. $a^0(\zeta) \neq a(\zeta)$. The intuition is that such a treatment does not eliminate all of the influence of latent demand – the decision-maker views both signals as informative and weighs them against one another, therefore the posterior belief lies between 0 and $E[h|h^L]$.[10] However, because signaling $h^T = 0$ moves actions toward the natural action it can be informative about the *direction* of latent demand. We also show that in an alternative formulation with non-independent signals, where participants perceive the demand treatments to contain the same information as latent demand but less noise, signaling $h^T = 0$ does elicit the natural action.

In sum, demanding the natural action is not guaranteed to obtain bounds that contain the natural action, while a pair of sufficiently strong positive and negative demand treatments does.

2.4 Inference

The theory tells us how to measure bounds on actions and treatment effects. Since these objects can be estimated experimentally, we may also wish to perform inference on the bounds themselves, or on the underlying parameters contained by the bounds. In Appendix C.7 we show how to do this, following Imbens and Manski (2004). We also provide a Stata package that allows calculation of demand-robust confidence intervals for mean behavior and treatment effects.

[10] One interpretation of latent influences in this model is "implicit" influence – the subject is not fully aware of the influence of latent demand and therefore unable to fully ignore it.

3 Sample and experimental design

We conducted six experiments in total to test the method and to provide estimates of demand sensitivity on a wide range of experimental tasks. We conduct all of our experiments online, primarily because the very large number of treatments would be infeasible to implement in the laboratory.

We purposefully designed the experiments to maximize comparability. For all experiments except the effort task, choice sets are similar (they can be expressed as real numbers from 0 to 1), we pay the same show-up fee, use a similar subject pool and mode of collection (online experiments with MTurk and online panel), the response mode (sliders) and as similar stake sizes as possible.[11]

In Table A.3 we summarize the key design features of each experiment. We describe the sample, which games were used, which demand treatments we employed, and whether choices involved real stakes or were hypothetical. More details on the experimental designs and the exact experimental instructions can be found in the pre-analysis plans, as well as the experimental instructions in the online appendix.

3.1 Participant populations

We conducted in five experiments with approximately 15,000 participants on Amazon Mechanical Turk (MTurk), and one exepriment with 3,000 respondents using a representative online panel of the US population. MTurk is an online labor marketplace that is frequently used by academics to recruit participants for experiments (Paolacci and Chandler, 2014; de Quidt, 2017; Bordalo et al., 2016; Kuziemko et al., 2015; DellaVigna and Pope, 2016a). It is attractvie because it offers researchers a large and diverse pool

[11]For the effort task, we replicated the design of DellaVigna and Pope (2016b). The primary differences are a higher show-up fee and a different response mode (effort).

of workers that have been shown to be more attentive to instructions than college students (Hauser and Schwarz, 2016).

To participate in our MTurk experiment, people had to live in the U.S, have an overall approval rating of more than 95%, and have completed more than 500 tasks on MTurk. We excluded prior participants when recruiting for experiments 2 and 3. However, in subsequent experiments, we had essentially exhausted the active participant pool, so to avoid undue delays in recruitment we allowed prior participants to take part.[12]

Most workers on MTurk are experienced in taking surveys, which is a potential threat to the external validity of our results (Chandler et al., 2014). We therefore conducted one additional experiment using a representative online sample, whose participants are less experienced with social science experiments.[13] This sample of 3,000 respondents is representative of the US population in terms of region, age, income and gender.

3.2 Summary statistics

The experiments were run between May and September 2016.[14] In Tables A.42 to A.47, we provide details on the sample characteristics of our respondents from both MTurk and the representative sample. In addition, in Tables A.35 to A.40, we present the pre-specified balance tables for all of these experiments. Table A.45 highlights that respondents from the online panel are representative of the US population by gender, income, age and region and in terms of other observables, such as education and race. Attrition was low, with less than 2 percent of participants dropping out of

[12]Our results are virtually unchanged by the exclusion of respondents that completed more than one of our experiments.

[13]We collect data on this sample through an online survey in collaboration with the market research company, Research Now. This provider has been used in previous research, for example by Almås et al. (2016).

[14]The pre-analysis plans were posted on the on the social science registry and can be found here: `https://www.socialscienceregistry.org/trials/1248`

our experiments on average. Importantly, there was no differential attrition across the different demand treatment arms. Tables A.17 and A.18 summarize attrition rates at the game level for the strong and weak demand experiments, respectively.

4 Applying the method

4.1 Bounding natural actions

Our first set of experiments attempts to measure the upper and lower bound of behavior using our strong demand treatments which explicitly tell participants they will "do us a favor" by taking a higher or lower action than normal. Our respondents complete one of the following games: a dictator game, an investment game (to measure risk preferences), convex time budgets (to measure time preferences), a trust game (first mover or second mover), an ultimatum game (first mover or second mover), a lying game, a measure of ambiguity aversion, and a real effort task.

As an example, in the dictator game, participants in the positive demand condition are given the following message: "You will do us a favor if you give more to the other participant than you normally would".[15] Participants in the negative demand condition receive the following message: "You will do us a favor if you give less to the other participant than you normally would." In Table A.4, we describe the key design features and demand instructions for each game.

For a subset of games[16], we also measure behavior in a no-demand

[15]Our instructions are close to the ones used by Binmore et al. (1985): "You will be doing us a favor if you simply set out to maximize your winnings". Deutsch et al. (1967) employ a similar approach, telling subjects "I want you to to earn as much money as you can regardless of how much the other earns". Such instructions have been criticized precisely because they risk inducing experimenter demand (Thaler, 1988; Zizzo, 2010).

[16]The dictator game, convex time budgets, the investment game and the two real effort tasks

condition in which participants receive no demand manipulation, to test Assumption 1, monotone demand treatment effects. In addition, in the dictator game, convex time budgets, and the investment game, half of our respondents made choices for real stakes, while half made hypothetical choices for the same stake sizes. For the remaining games, all choices are incentivized.

Table 1 and Figures 1 and 2 summarize the response to the strong demand treatments for each of the games, restricting the sample to MTurk respondents and incentivized choices. In Panel A of Table 1, we show the unconditional mean actions in the different demand treatment arms.[17] Our objects of interest are mean behavior in the positive demand condition, $a^+(\zeta)$, mean behavior in the negative demand condition, $a^-(\zeta)$, and mean behavior in the no-demand condition, $a^L(\zeta)$.

In Panel B of Table 1, we display our sensitivity measure $(a^+(\zeta)-a^-(\zeta))$ for each of the 11 games, in both raw and z-scored units.[18]

Behavior is responsive to our strong demand treatments across tasks, and sensitivity is significantly different from zero in all tasks. Sensitivity is particularly high in the dictator game, for second movers in the trust game and ultimatum games, and for unincentivized effort.

In Panel C of Table 1, we examine the monotone demand treatment effects assumption: $a^+(\zeta) \geq a^L(\zeta) \geq a^-(\zeta)$. We estimate the following equation, in which POS_i takes value one for people in the positive demand condition and value zero otherwise, and where NEG_i takes value one for people in the negative demand condition and value zero otherwise:

$$ZY_i = \pi_0 + \pi_1 POS_i + \pi_2 NEG_i + \varepsilon_i \tag{8}$$

[17] For most of the games the action set lies between 0 and 1; if not we rescale the action space to the $[0, 1]$ interval.

[18] We z-score our outcome variables at the game level, using the mean and s.d. for the negative-demand group (Kling et al., 2007).

As can be seen in Panel C of Table 1, we find support for the assumption. In all cases the positive demand treatment increased actions and the negative treatment decreased them. In most cases the differences are statistically significant.

[Insert Table 1 and Figures 1 2]

4.2 Bounding treatment effects

In our real effort experiment, which replicates a subset of the treatments in DellaVigna and Pope (2016b), participants earned points by alternately pressing two keyboard buttons for 10 minutes. In one treatment arm respondents were told that their score "will not affect [their] payment," while in the second they received one cent per 100 points. For some participants we added to these our demand treatments, telling workers "you will do us a favor if you work harder/less hard than you normally would." We can apply our method to estimate bounds on the treatment effect of performance pay on effort provision.[19]

Panel A in Figure 3 and Panel A in Table 2 summarize the sensitivity of effort to our demand treatments and how this sensitivity depends on incentives. We find that individuals who receive no bonus were substantially more sensitive to our demand treatments.

Panel B in Figure 3 and Panel B in Table 2 display the conventional treatment effect ($a^L(1) - a^L(0)$, where "1" and "0" correspond to the reward per 100 points), the upper bound of the treatment effect ($a^+(1) - a^-(0)$), and the lower bound of the treatment effect ($a^-(1) - a^+(0)$).

The bounds we estimate are quite wide, ranging from 0.25 to 1.35 standardized units. However, the lower bound (0.25) is statistically significantly

[19] In addition to those main treatment arms, 250 additional individuals completed the "no demand condition" for a reward of 4 cents per 100 points they score. This treatment arm is used in the structural estimation below.

different from zero. This implies that even if behavior in the unincentivized condition is biased by extreme negative latent demand (i.e. $a(0) = a^+(0)$), while behavior in the incentivized condition is strongly positively biased (i.e. $a(1) = a^-(1)$), we can still support the conclusion that incentives increase effort.

[Insert Table 2 and Figure 3]

4.3 Structural estimates

Under additional parametric and identifying assumptions, our demand treatments permit structural estimation of demand-free model parameters (v), as well as ϕ and the latent demand beliefs. Knowing v allows the researcher to make predictions about behavior in demand-free environments, knowing ϕ allows them to quantify the importance of experimenter demand relative to v, while measuring beliefs can enable them to diagnose and eliminate the sources of latent demand effects. We illustrate how structural estimation can be performed using the real effort experiment. Because demand effects can be easily incorporated in the model of DellaVigna and Pope (2016b) (DP), we can exactly follow their approach to structural estimation.

In their experiment, DP estimate the following utility function (expressed in our notation):

$$v(a) = (s + \zeta)a - c(a) \tag{9}$$

The action a is effort, measured in points on the task, $c(a)$ is a cost of effort function, ζ is a piece rate and s is an intrinsic motivation parameter – workers may work for no pay because they enjoy the task. DP solve the first order condition and estimate the model parameters using nonlinear least squares (NLLS).[20]

Adding demand to this utility function gives:

$$U(a, \zeta) = (s + \zeta + \phi(\zeta)E[h|h^T, h^L(\zeta)])a - c(a) \tag{10}$$

with corresponding first-order condition

$$s + \zeta + \phi(\zeta)E[h|h^T, h^L(\zeta)] - c'(a^*(\zeta)) = 0 \tag{11}$$

DP consider two alternative forms for c, a power function $c(a) = ka^{1+\gamma}/(1+\gamma)$, yielding optimal effort equal to:

$$a^*(\zeta) = \left(\frac{s + \zeta + \phi(\zeta)E[h|h^T, h^L(\zeta)]}{k}\right)^{\frac{1}{\gamma}} \tag{12}$$

and an exponential form $c(a) = k\exp(\gamma a)/\gamma$ with corresponding effort level:

$$a^*(\zeta) = \frac{1}{\gamma}\log\left(\frac{s + \zeta + \phi(\zeta)E[h|h^T, h^L(\zeta)]}{k}\right) \tag{13}$$

We have seven treatment groups in total: neutral treatments with piece rates equal to 0 cents, 1 cent, and 4 cents per 100 points on the task, and positive and negative strong demand treatments in the 0 and 1 cent groups.[21] However, we have 13 parameters in total: s, k, γ, $\phi(0)$, $\phi(1)$, $\phi(4)$, $p^L(0)$, $p^L(1)$, $p^L(4)$, $h^L(0)$, $h^L(1)$, $h^L(4)$ and p^T.[22] We therefore need to impose some further restrictions.

First we assume that ϕ is fixed: $\phi(0) = \phi(1) = \phi(4) = \phi$. In other words, varying incentives do not change the participants' desire to please

[20] They also employ a minimum distance estimation procedure, we stick to NLLS for brevity.

[21] We also collected data using weak demand treatments, but we do not use it in this analysis a) because it was collected in a separate experiment and b) because as we explain below, for estimation we need to impose the parameter restriction $p^T = 1$, which we do not believe is satisfied in the weak treatments.

[22] In principle p^T might also vary with ζ and h^T. Our model presented above rules that out by assumption.

the experimenter. Second, we assume $p^T = 1$. By assumption this is not justified for our weak demand treatments, so we focus on the strong treatments where the assumption is more plausible. This assumption implies that $E[h|h^T, h^L] = h^T$. We are therefore able to identify ϕ, s, γ and k just using the four demand treatment groups. Third, since $E[h|h^L(\zeta)] = p^L(\zeta)h^L(\zeta) \in [-1, 1]$, we can treat it as a single parameter whose sign identifies h^L and whose magnitude identifies $p^L(\zeta)$. We are left with seven parameters, s, k, γ, ϕ, $p^L(0)h^L(0)$, $p^L(1)h^L(1)$, $p^L(4)h^L(4)$, and are therefore exactly identified. We additionally estimate a specification in which we restrict latent demand to depend only on whether monetary incentives are present, i.e. $p^L(1)h^L(1) = p^L(4)h^L(4)$, in which case we are overidentified.

While we estimate the same model as DP, the identification comes from a different source. Under the assumption of no latent demand (as in DP), s, γ and k are identified from the three neutral treatments. When latent demand is present, the model parameters are identified from the demand treatments and the neutral treatments identify the latent demand beliefs. This also means that for our purposes the neutral treatment with four cent incentives is not necessary for identification of the core parameters.

We follow DP in estimating equation (12) in logs and equation (13) in levels, using nonlinear least squares.[23] Estimation results are presented in Table 3. Columns 1–3 correspond to the power cost function and columns 4–6 to the exponential cost function. In each case we first mirror DP by estimating s, γ and k using only the neutral treatments, assuming that there is no latent demand.[24] Second, we include all treatment groups and

[23]Since the piece rates are per 100 points, we follow DP in rounding scores to the nearest 100. See Appendix D for further details of the estimation.

[24]The parameter estimates we obtain are quite different from those of DP. One possible explanation is that DP estimate their main specification from 0, 1 and 10 cent treatments while we use 0, 1 and 4 cents, which may discipline the curvature of the effort cost function less. Additionally, while we like they recruited our participants on MTurk, our experiments were conducted some time after theirs so the subject pool may have changed somewhat.

impose that latent demand depends only on whether monetary incentives are present. Third, we allow latent demand to differ across all three incentive levels. Coefficients s and ϕ are measured in cents per 100 points. Therefore, $s = 1$ is interpreted as intrinsic demand playing an equivalent role to an incentive of 1 cent per 100 points, while $\phi = 1$ means that a worker who is certain the experimenter wants high effort works as if her incentives were increased by 1 cent per 100 points, relative to someone who does not know the experimenter's wishes.

Our first finding is an important role for experimenter demand. Our estimates of s and ϕ are quite large and of similar magnitude in all specifications, taking values equivalent to 0.2–0.5 cents per 100 points. Assuming a value of ϕ of around 0.25 would imply that switching from extreme negative to extreme positive demand (a change in $E[h|h^L]$ from -1 to 1) increases effort by as much as increasing the incentive by 0.5 cents per 100 points. Our estimates of $E[h|h^L]$ are mostly negative, consistent with latent demand decreasing effort (though in the exponential cost case we estimate a positive value in the 1 cent treatment). However, the estimates are noisy and typically not significantly different from zero. We also estimate that in the 4 cent treatment, $E[h|h^L(4)] \approx -6.5$, which contradicts the theory (beliefs should lie in $[-1, 1]$), though we note that -1 lies well within the 95% confidence interval. We do not have demand treatment groups who were paid 4 cents, so the effort cost function is extrapolated out of sample to this group, which may help explain the poor fit of the model.

Our second finding is that allowing for demand is quantitatively important for other parameter estimates. This is most noticeable when comparing the estimates of s when we do and do not allow for demand effects; the estimates are an order of magnitude smaller in the latter case (columns 1 and 4). Our estimates imply a negative latent demand effect which is instead attributed to low intrinsic motivation.

Finally, the structural estimates enable us to go beyond bounding to back out predicted demand-free behavior. To do this we plug our parameter estimates back into the first order conditions, fixing $E[h|h^L] = 0$. Results are presented in table A.6. Since most of our estimated latent demand effects are negative, predicted demand-free effort is usually higher, than observed effort, sometimes considerably so.

[Insert Table 3]

4.4 Tighter bounds with weak demand treatments

While the bounds we obtain using the strong demand treatments satisfy the monotone demand treatment effects assumption, they are wide, meaning that we cannot rule out many possible values for $a(\zeta)$. We therefore conduct additional experiments[25] to obtain less conservative bounds, at the cost of a less plausible bounding assumption.

Our weak demand treatments take a similar form to the strong treatments, but rather than explicitly reporting an objective ("you will do us a favor") we simply reveal an experimental hypothesis to the participants, without demanding that they act to confirm it. For example, in the investment game, participants were told that "We expect that participants who are shown these instructions will invest more/less in the project than they normally would," with the phrasing modified accordingly for other tasks. Table A.5 describes design features and the demand instructions for each game.

As before, for a subset of games (the dictator and investment games) we collect data without the demand treatment to test for monotone demand treatment effects. In addition, for the dictator game and the investment

[25] We employed weak demand treatments in experiments 2, 4, 5 and 6. A more detailed description can be found in Table A.3

game, half of our respondents' choices are incentivized, and half hypothetical. For all other tasks all choices are incentivized.

We present results from the MTurk experiments with weak demand treatments and incentivized choices. In Panel A of Table 4 we display the unconditional means in the different demand treatment arms. In Figure 2 we plot these values with confidence intervals, while in Figure 1 we display the sensitivities. Our objects of interest are mean behavior in the positive demand condition, $a^+(\zeta)$, mean behavior in the negative demand condition, $a^-(\zeta)$, and mean behavior in the no-demand condition, $a^L(\zeta)$.

Panel B of Table 4 displays the sensitivities by game, $(a^+(\zeta) - a^-(\zeta))$.[26] For convex time budgets, the effort tasks, the lying task, and the trust game first mover, we find sensitivities to weak demand below 0.1 standard deviations.[27] We find stronger responses (between 0.2–0.25 standard deviations) for the dictator game, the ultimatum game second mover, and the trust game second mover. Sensitivity in the investment game under risk and uncertainty, as well as for the ultimatum game first mover, is approximately 0.17 standard deviations.

In Panel C of Table 4 we examine the monotone demand treatment effect assumption, finding that most demand treatments have the correct sign and are significant. We estimate a small negative effect of the positive demand treatment in the investment game, and a small positive effect of the negative treatment in the dictator game, though neither estimate is statistically significant.

[Insert Table 4]

[26]We z-score our outcome variables at the paradigm-incentive level, using the mean and s.d. for the negative-demand group (Kling et al., 2007).

[27]Our minimum detectable effect sizes range from between .2 to .34 standard deviations.

4.5 Confidence intervals

We compute confidence intervals for (a) the bounds themselves, and (b) for the parameters contained by those bounds (an action or treatment effect), following Imbens and Manski (2004). Confidence intervals for the parameter are slightly tighter than for the bounds, which reflects the fact that the parameter cannot lie at both bounds simultaneously.[28] In Table A.7 we present confidence intervals for all 11 games, and in Table A.8 we present confidence intervals for the treatment effect in the effort experiment. Details on how confidence intervals are computed are given in Appendix C.7.

4.6 Moderators of sensitivity to demand

Does sensitivity to demand treatments vary with design and participant characteristics? In this section, we examine heterogeneous responses to our strong and weak demand treatments by whether choices are incentivized or hypothetical, gender, attentiveness, and participant pool (MTurk vs. representative online panel). Whether or not this heterogeneity can be interpreted as informative about differences in underlying latent demand (e.g., whether greater sensitivity among one gender reflects a greater influence of latent demand for that gender) depends upon whether Monotone Sensitivity holds for the subset of environments under consideration, i.e. whether they belong to the same comparison class. We show in Appendix C.4 that variation in incentives, attention and the preference for pleasing the experimenter, ϕ (which may differ by gender or participant pool) plausibly form valid bases for comparison classes.

[28]But for a coverage correction derived by Imbens and Manski, the $1-\alpha$ confidence interval on the parameter corresponds to the $1-2\alpha$ interval on the bounds.

4.6.1 Incentivized vs. hypothetical choices

In the dictator game, in the investment game, and in the convex time budgets, we randomly assigned participants to make either hypothetical or incentivized choices. In what follows, we test whether making an incentivized rather than a hypothetical choice affects participants' response to our strong demand treatments. To do so, we regress our standardized outcome variables[29], pooled across games, on a demand treatment indicator, POS_i, taking value one for people in the positive demand condition and value zero for people in the negative demand condition; an indicator M_i, taking value one for the incentive condition; and the interaction between M_i and POS_i:

$$ZY_i = \beta_0 + \beta_1 POS_i + \beta_2 M_i \times POS_i + \beta_3 M_i + \varepsilon_i \tag{14}$$

Results are presented in Figure 4 and Panel B of Table A.1. Surprisingly, participants making hypothetical or incentivized choices responded very similarly to experimenter demand.[30,31] It is surprising that sensitivity did not seem to respond to incentives, since deviations from the natural action are presumably less costly in hypothetical choice. Possibly this reflects that even our incentivized choices were relatively low-stakes and that we would see a difference at higher stakes. Our results relate to previous work examining the effects of incentives on behavior in the lab (Camerer et al., 1999).

However, we note that in the effort experiment we do see lower sen-

[29] We standardize the outcome variable at the game-incentive level.

[30] We included a manipulation check showing that respondents making incentivized choices were more likely to believe that the tasks involved real money compared to respondents completing hypothetical choices. See table A.24.

[31] As we discuss in section 2, we can meaningfully compare effect sizes across the respondents making incentivized choices rather than hypothetical choices if the monotone sensitivity assumption holds. In Appendix C.4 we show that this is the case if utility is additively separable or multiplicatively separable in incentives.

sitivity, measured either in points or standardized units, when effort is incentivized than when unincentivized (note that unincentivized effort is conceptually different from hypothetical choice).[32]

[Insert Figure A.1]

4.6.2 Gender and attention

We measure participant gender and attentiveness in all tasks. We define a participant as attentive if they passed an attention screener at the beginning of the task.[33] To examine heterogeneous responses to our strong demand treatments by these variables, we estimate the following equation:

$$ZY_i = \beta_0 + \beta_1 POS_i + \beta_2 Interaction_i \times POS_i + \beta_3 Interaction_i + \varepsilon_i \quad (15)$$

where $Interaction_i$ is the dimension of heterogeneity of interest.

First, we test whether the normalized[34] sensitivity to experimenter demand differs for men and women using a dummy, $Male_i$, taking value one for males. As can be seen in Table A.1 and Figure A.2 , we find that females respond more strongly to the strong demand treatments than males. [35]

[Insert Figure A.2]

As we show in Appendix C.4, if the difference in sensitivity is driven by differences in willingness to please the experimenter, then it is indicative

[32] In the model in section 4.3 effort is additively separable in incentives, satisfying the condition for Monotone Sensitivity in Appendix C.4.

[33] The screener presents participants with a paragraph of text that appears to direct them to select their preferred online news sources from a list, but concealed in the text is an instruction to ignore this and choose two specific options. The assumption is that attentive respondents read the question and follow the concealed instruction, inattentive respondents do not.

[34] We normalize the outcome variable at the game-gender level.

[35] Our finding is related to the large literature on gender differences in preferences (Croson and Gneezy, 2009).

of stronger latent demand effects for females. However males and females might also hold different beliefs about the experimental objective, in which case Monotone Sensitivity would not hold.[36]

We also examine whether respondents who did not pass an attention screener at the beginning of our experiments respond to our demand treatments differently from respondents who passed. Turning to attention, Table A.1 and Figure A.3, find higher sensitivity among attentive MTurkers, but this effect is not significant and noisily measured as only 10 percent of our MTurk sample were inattentive.[37]

[Insert Figure A.3 and Table A.1]

In the representative online panel, however, we have enough variation to examine heterogeneous effects by attention with sufficient statistical power. As can be seen in Table A.29 we find evidence that participants who passed the screener were significantly more sensitive to the demand treatments than those who did not. The estimated difference between attentive and inattentive respondents from the representative sample is quite similar to that of MTurkers. If we pool the data from MTurk and the representative online panel, we find significantly different sensitivities to our demand treatments by attention. As discussed in the Appendix C.4, variation in sensitivity generated by inattention to experimenter demand satisfies the monotone sensitivity assumption.[38] [39]

[36]Using the belief measure described in section 4.7 we do not find any evidence that males and females in the no-demand condition hold different beliefs, nor do they update their beliefs differently in response to the demand treatments.

[37]We normalize the outcome at the game-attention level.

[38]Consistent with that model, our beliefs data suggest that attentive respondents update their beliefs about the experimental objective more strongly in response to our demand treatments.

[39]We also examine heterogeneous sensitivities to our strong demand treatments along two additional dimensions that we had not pre-specified, education and prior experience with experiments measured by the number of HITs previously completed on MTurk. We find no evidence that more experienced or more educated respondents react more to our demand treatments. These results are summarized in Table A.11.

We also examine heterogeneous responses to the weak demand treatments, though we have less power to detect differences because of the lower overall sensitivity to these treatments. We find no significant differences in sensitivity by incentives, attention, gender, education or experience. These results are summarized in Tables A.2 and A.12.

4.6.3 MTurk vs. representative online panel

Some experimental social scientists are concerned that MTurk workers are experienced research participants and may behave differently to a more representative participant pool. Moreover, MTurkers need to maintain a high "approval" rating and may therefore be especially motivated to please the researcher [40]

To address such concerns, and to test an additional dimension of heterogeneity, we replicated the MTurk dictator game and investment game experiments with a representative online survey panel, whose participants are less experienced experimental subjects. We randomly assigned respondents to either a positive weak demand treatment, a negative weak demand treatment, a positive strong demand treatment, a negative strong demand treatment or no demand treatment. All of our respondents' choices in this experiment were incentivized and the stake size was the same as in the MTurk experiment.

In Table A.1 and Figure A.4 we test for differences in sensitivity in both the pooled dictator and investment games, and for each game separately. Representative panel subjects responded more strongly in the risk game and less strongly in the dictator game (significant at 10%); the pooled test finds a small and non-significant difference in sensitivity. It seems that MTurkers are not especially more or less susceptible to experimenter

[40] Recruiters on MTurk have the option to reject unsatisfactory work, and recruiters can screen out workers with high rejection rates. However we believe it is well-known that researchers rarely reject work; we never did. See Berinsky et al. (2014).

demand.[41]

4.7 Beliefs

In each experiment, after participants had completed the relevant task, we collected simple, unincentivized belief data. Specifically, we asked two questions. First, we asked "What do you think is the result that the researchers of this study want to find?" and second "What do you think was the hypothesis of this research study?" Responses were binary, subjects could respond that they thought the objective/hypothesis was either a high or low action. The purpose of these belief measures was as a manipulation check, to ensure that subjects' beliefs responded as expected to the demand treatments.[42] Additionally any belief measure is likely to be subject to its own latent demand bias, so we caution against overinterpreting the data.

We think there are two natural interpretations of the belief responses. If participants report a high belief if their posterior $E[h|h^T, h^L]$ is positive and a low belief if negative, then the average response tells us the fraction of participants with high beliefs but not the average belief. If instead participants randomize according to the strength of their beliefs (reporting $h = 1$ with probability $Pr[h = 1|h^T, h^L]$) then the average response equals the average belief.

Results are presented in Tables A.13, A.14, A.15 and A.16 and confirm

[41]Respondents from MTurk and the representative online panel might exhibit different willingness to please the experimenter, but they could also have different beliefs or attentiveness. In line with the above finding that respondents from the online panel are less attentive, they also updated their beliefs about the experimental objective less than MTurkers. Focusing on the subsample of attentive respondents from MTurk and the representative online panel, we find that respondents from the representative online panel update their beliefs more. This suggests that some of the variation across groups is driven by differences in beliefs and hence monotone sensitivity may not hold.

[42]While in principle once could collect richer belief measures and incentivize responses we opted for this simple approach, in part because asking for fine-grained beliefs about *our* motivations seemed quite unnatural and because there was no objective truth against which to score. Techniques do exist for belief scoring without an objective truth, e.g. Prelec (2004).

that our treatments moved average responses in the correct direction. Interestingly the magnitude of the shift in both belief measures is similar for the strong and weak treatments, and in some cases the weak treatments shifted beliefs by more than the strong treatments (note that not all strong and weak treatments were conducted in the same experiments). For example, in the dictator game, 65% of the strong positive demand treatment group reported that the researchers "want to find that on average people give a large share of the $1 to the other person," (alternative, "small"). Under the strong negative demand treatment only 24% so responded. The corresponding figures are 54% and 23% for the weak treatment.

The fact that reported beliefs were very similar in the strong and weak treatments suggests that modeling demand treatment strength by informativeness of the contained signal may be missing something. One interpretation is that while both treatments are equally informative about the *direction* of experimenter demand, the strong treatments make the experimental objective appear more important or more salient, i.e. ϕ is larger under the strong treatments, leading to wider bounds. We discuss in section 2.3.6 and derive in the appendix an extension to the model in which respondents learn about both h and ϕ. Bounding still relies on $p^T \geq p^L$ but weak treatments can obtain tighter bounds by signaling that ϕ is small.

A final use of the belief data is as an alternative (not pre-specified) measure of attentiveness: we can classify as attentive those participants reporting the "correct" belief about the experimental objective in response to our demand treatments. A.10 shows that sensitivity to our strong demand treatments is high–around 1 standard deviation–among participants considered attentive by this measure. Sensitivity for inattentive subjects is close to zero and never significant.

5 Conclusion

We propose a technique for assessing the robustness of behavior to experimenter demand. We deliberately induce demand in a structured way to measure its influence on behavior and to construct bounds on demand-free behavior and treatment effects. We formalize the intuition behind explicit demand treatments with a simple model in which participants in an experiment form beliefs about the experimental objective and receive utility from conforming to it. Bounds are obtained by intentionally manipulating those beliefs.

We find that behavior in eleven canonical economic games is quite sensitive to our strong demand manipulations that explicitly signal an experimental objective, generating bounds of up to 1 standard deviation in width. Much tighter bounds are obtained using weak demand treatments in which we signal only an experimental hypothesis. We expect that the latter are a more realistic measure of the magnitude of demand effects in the typical experiments.

We also show how to analyze demand effects structurally, following the approach of (DellaVigna and Pope, 2016b). Our estimates suggest a utility from pleasing the experimenter worth approximately one quarter the size of the monetary incentive. We leverage the structural model to extract predictions for demand-free behavior, or "natural actions" in our terminology.

Finally, we find that females respond more to our strong demand treatments than males, but no significant heterogeneous sensitivity to demand by whether choices are incentivized, education, prior experience or the subject pool. We find some evidence that more attentive participants conform more to our strong demand treatments.

Future work might employ similar treatments to study how one can lower demand in experiments. Researchers could examine what features of the environment, the experimenter, and the mode of data collection (e.g. online vs. laboratory) influence demand effects.

References

Almås, Ingvild, Alexander Cappelen, and Bertil Tungodden, "Cutthroat Capitalism versus Cuddly Socialism: Are Americans more Meritocratic and Efficiency-seeking than Scandinavians?," *Working Paper*, 2016.

Altonji, Joseph G, Todd E Elder, and Christopher R Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 2005, *113* (1), 151–184.

Angrist, Joshua and Guido Imbens, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–476.

Bardsley, Nicholas, "Dictator Game Giving: Altruism or Artefact?," *Experimental Economics*, 2008, *11* (2), 122–133.

Barmettler, Franziska, Ernst Fehr, and Christian Zehnder, "Big Experimenter is Watching you! Anonymity and Prosocial Behavior in the Laboratory," *Games and Economic Behavior*, 2012, *75* (1), 17–34.

Berinsky, Adam J, Michele F Margolis, and Michael W Sances, "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys," *American Journal of Political Science*, 2014, *58* (3), 739–753.

Binmore, Ken, Avner Shaked, and John Sutton, "Testing Noncooperative Bargaining Theory: A Preliminary Study," *The American Economic Review*, 1985, *75* (5), 1178–1180.

Bó, Ernesto Dal and Pedro Dal Bó, ""Do the Right Thing:" The Effects of Moral Suasion on Cooperation," *Journal of Public Economics*, 2014, *117*, 28–38.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, "Stereotypes," *The Quarterly Journal of Economics*, 2016, *131* (4), 1753–1794.

Camerer, Colin F, Robin M Hogarth, David V Budescu, and Catherine Eckel, "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-production framework," *Journal of Risk and Uncertainty*, 1999, pp. 7–48.

Chandler, Jesse, Pam Mueller, and Gabriele Paolacci, "Nonnaïveté among Amazon Mechanical Turk Workers: Consequences and Solutions for Behavioral Researchers," *Behavior Research Methods*, 2014, *46* (1), 112–130.

Chassang, Sylvain, Gerard Padró i Miquel, and Erik Snowberg, "Selective trials: A principal-agent Approach to Randomized Controlled Experiments," *The American Economic Review*, 2012, *102* (4), 1279–1309.

Conley, Timothy G, Christian B Hansen, and Peter E Rossi, "Plausibly Exogenous," *Review of Economics and Statistics*, 2012, *94* (1), 260–272.

Croson, Rachel and Uri Gneezy, "Gender Differences in Preferences," *Journal of Economic literature*, 2009, *47* (2), 448–474.

de Quidt, Jonathan, "Your Loss is my Gain: a Recruitment Experiment with Framed Incentives," *Journal of the European Economic Association, forthcoming*, 2017.

DellaVigna, Stefano and Devin Pope, "Predicting Experimental Results: Who Knows What?," *National Bureau of Economic Research*, 2016.

_ and _ , "What Motivates Effort? Evidence and Expert Forecasts," *National Bureau of Economic Research*, 2016.

_ , John A List, and Ulrike Malmendier, "Testing for Altruism and Social Pressure in Charitable Giving," *The Quarterly Journal of Economics*, 2012, *127*, 1–56.

_ , _ , _ , and Gautam Rao, "Voting to Tell Others," *Review of Economic Studies, forthcoming*, 2017.

Deutsch, Morton, Yakov Epstein, Donnah Canavan, and Peter Gumpert, "Strategies of Inducing Cooperation: An Experimental Study," *Journal of Conflict Resolution*, 1967, *11* (3), 345–360.

Gerber, Alan S, Donald P Green, and Christopher W Larimer, "Social Pressure and Voter Turnout: Evidence from a Large-scale Field Experiment," *American Political Science Review*, 2008, *102* (01), 33–48.

Harrison, Glenn W and John A List, "Field Experiments," *Journal of Economic literature*, 2004, *42* (4), 1009–1055.

Hauser, David J and Norbert Schwarz, "Attentive Turkers: MTurk Participants Perform Better on Online Attention Checks Than do Subject Pool Participants," *Behavior research methods*, 2016, *48* (1), 400–7.

Hoffman, Elizabeth, Kevin A McCabe, and Vernon L Smith, "On Expectations and the Monetary Stakes in Ultimatum Games," *International Journal of Game Theory*, 1996, *25* (3), 289–301.

_ **, Kevin McCabe, Keith Shachat, and Vernon Smith**, "Preferences, Property Rights, and Anonymity in Bargaining Games," *Games and Economic behavior*, 1994, *7* (3), 346–380.

Imbens, Guido W and Charles F Manski, "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 2004, *72* (6), 1845–1857.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 2007, *75* (1), 83–119.

Kuziemko, Ilyana, Michael I Norton, Emmanuel Saez, and Stefanie Stantcheva, "How Elastic are Preferences for Redistribution? Evidence from Randomized Survey Experiments," *The American Economic Review*, 2015, *105* (4), 1478–1508.

Levitt, Steven D and John A List, "What do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?," *The Journal of Economic Perspectives*, 2007, *21* (2), 153–174.

List, John A, "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions," *Journal of Political Economy*, 2006, *114* (1), 1–37.

_ , "On the Interpretation of Giving in Dictator Games," *Journal of Political Economy*, 2007, *115* (3), 482–493.

_ **, Robert P Berrens, Alok K Bohara, and Joe Kerkvliet**, "Examining the Role of Social Isolation on Stated Preferences," *The American Economic Review*, 2004, *94* (3), 741–752.

Masclet, David, Charles Noussair, Steven Tucker, and Marie-Claire Villeval, "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism," *The American Economic Review*, 2003, *93* (1), 366–380.

Milgram, Stanley, "Behavioral Study of Obedience.," *The Journal of Abnormal and Social Psychology*, 1963, *67* (4), 371.

Nevo, Aviv and Adam M Rosen, "Identification with Imperfect Instruments," *Review of Economics and Statistics*, 2012, *94* (3), 659–671.

Orne, Martin T, "On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and their Implications.," *American psychologist*, 1962, *17* (11), 776.

Paolacci, Gabriele and Jesse Chandler, "Inside the Turk: Understanding Mechanical Turk as a Participant Pool," *Current Directions in Psychological Science*, 2014, *23* (3), 184–188.

Prelec, D., "A Bayesian Truth Serum for Subjective Data," *Science*, Oct 2004, *306* (5695), 462–466.

Rosenthal, Robert, *Experimenter Effects in Behavioral Research.*, Appleton-Century-Crofts, 1966.

Shmaya, Eran and Leeat Yariv, "Experiments on Decisions Under Uncertainty: A Theoretical Framework," *American Economic Review*, 2016, *106* (7), 1775–1801.

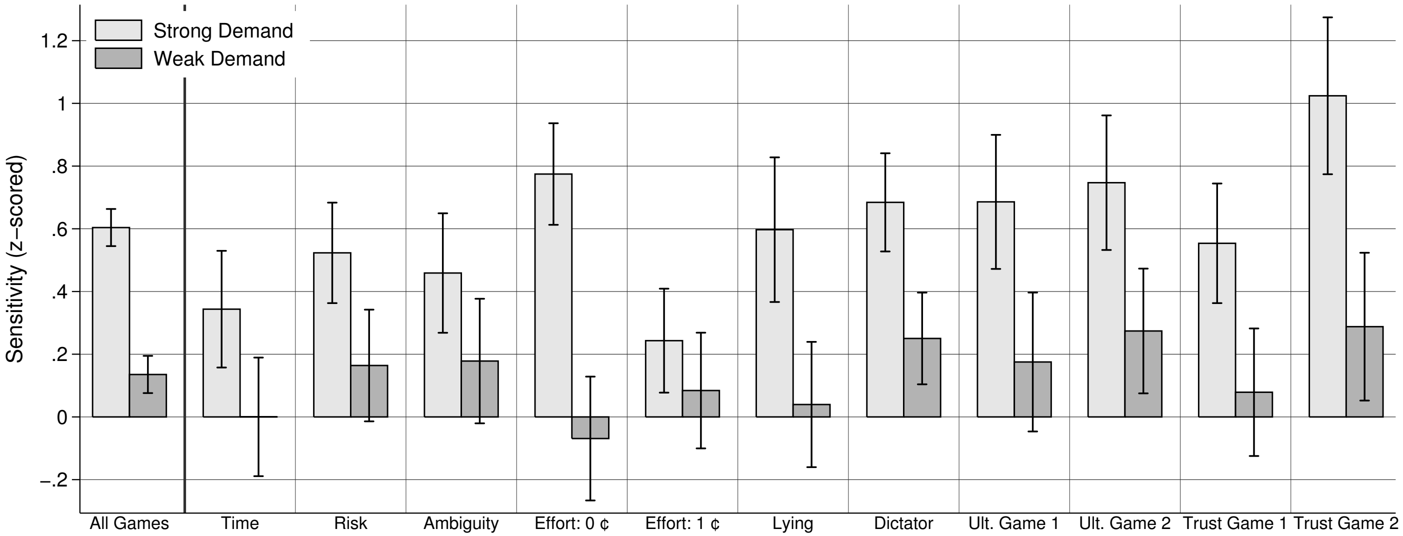
Thaler, Richard H, "Anomalies: The Ultimatum Game," *The Journal of Economic Perspectives*, 1988, *2* (4), 195–206.

Zizzo, Daniel John, "Experimenter Demand Effects in Economic Experiments," *Experimental Economics*, 2010, *13* (1), 75–98.

— **and Piers Fleming**, "Can Experimental Measures of Sensitivity to Social Pressure Predict Public Good Contribution?," *Economics Letters*, 2011, *111* (3), 239–242.

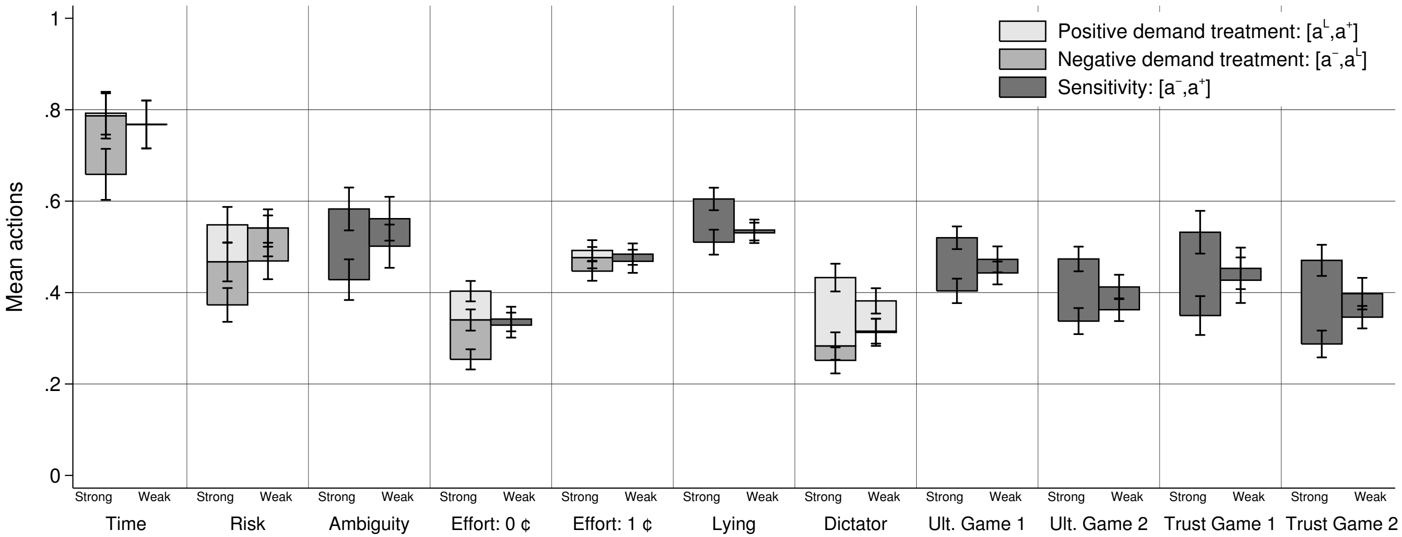
6 Main Figures and Tables

Figure 1: Effect of strong demand on mean behavior



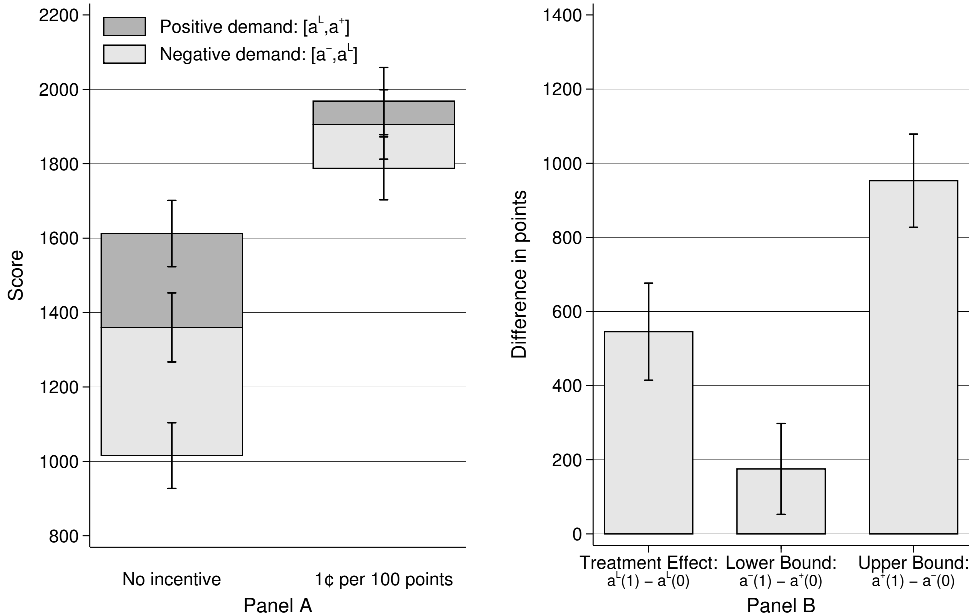
Notes: This figure utilizes data from all MTurk experiments with strong and weak demand treatments using real stakes. We present the z-scored sensitivity of behavior to our demand treatments, i.e. the normalized difference in behavior in the positive and negative demand condition. In our strong demand treatments we reveal the experimental objective to our respondents, while in the weak demand treatment we reveal the experiment hypothesis.

Figure 2: Effect of weak demand on mean behavior



Notes: This figure utilizes data from all MTurk experiments with strong and weak demand treatments using real stakes. This figure displays the response to our strong and weak demand treatments across 11 standard preference measures. We display the means and the 95 percent confidence intervals for the average behavior in the positive demand treatment arm, the negative demand treatment arm as well as the "no-demand" treatment arm. In the strong demand treatments we reveal the experimental hypothesis to our respondents, while in the weak demand treatment we reveal an experimental hypothesis to our respondents.

Figure 3: Effect of strong demand on treatment effect



Notes: This figure utilizes MTurk data from the effort experiment using the strong demand treatments (experiment 3). Panel A displays mean behavior in the different demand treatment arms disaggregated by the incentive condition alongside with the 95 percent confidence interval. Panel B displays the upper and lower bounds of treatment effect estimates as well as the confidence intervals for our treatment effect estimates. In these demand treatments we reveal the experimental objective to our respondents.

Table 1: Response to strong demand treatments across all incentivized games

	Time	Risk	Ambiguity Aversion	Effort 0 cent bonus	Effort 1 cent bonus	Lying	Dictator Game	Ultimatum Game 1	Ultimatum Game 2	Trust Game 1	Trust Game 2
Panel A: Unconditional Means											
Positive demand	0.792	0.548	0.583	0.403	0.492	0.605	0.433	0.520	0.474	0.532	0.470
	(0.024)	(0.020)	(0.024)	(0.011)	(0.011)	(0.012)	(0.015)	(0.013)	(0.014)	(0.024)	(0.017)
No demand	0.786	0.467		0.340	0.476		0.283				
	(0.025)	(0.022)		(0.012)	(0.012)		(0.015)				
Negative demand	0.659	0.373	0.428	0.254	0.447	0.510	0.252	0.404	0.338	0.350	0.288
	(0.028)	(0.019)	(0.023)	(0.011)	(0.011)	(0.014)	(0.014)	(0.014)	(0.014)	(0.022)	(0.015)
Panel B: Sensitivity (positive - negative)											
Raw data	0.134***	0.175***	0.155***	0.149***	0.045***	0.095***	0.181***	0.116***	0.136***	0.182***	0.183***
	(0.037)	(0.027)	(0.033)	(0.016)	(0.016)	(0.019)	(0.021)	(0.018)	(0.020)	(0.032)	(0.023)
Z-score	0.344***	0.523***	0.459***	0.775***	0.243***	0.597***	0.684***	0.686***	0.747***	0.554***	1.024***
	(0.095)	(0.082)	(0.097)	(0.083)	(0.085)	(0.118)	(0.080)	(0.109)	(0.109)	(0.097)	(0.128)
	[0.001]	[0.001]		[0.001]	[0.012]		[0.001]				
Panel C: Monotonicity											
Positive - Neutral (z-score)	0.015	0.242***		0.328***	0.085		0.566***				
	(0.089)	(0.088)		(0.085)	(0.089)		(0.082)				
	[0.404]	[0.002]		[0.001]	[0.128]		[0.001]				
Negative - Neutral (z-score)	-0.328***	-0.281***		-0.447***	-0.159*		-0.118				
	(0.097)	(0.086)		(0.084)	(0.086)		(0.079)				
	[0.001]	[0.001]		[0.001]	[0.070]		[0.047]				
Observations	730	730	404	735	717	366	773	409	425	383	373

Notes: This table uses data from all MTurk experiments with strong demand treatments using real stakes. In Panel A we display the unconditional means and standard errors of those means in the positive, negative and no-demand treatment arms respectively. In Panel B we present the raw and z-scored sensitivity of behavior to our demand treatments. In Panel C we display the sensitivity of behavior in the positive and negative demand condition compared to the no-demand condition. In the demand treatments we reveal the experimental objective to our respondents. Robust standard errors are in parentheses. False-discovery reate adjusted p-values are in brackets. The p-value of an F-test which tests for differences in response to demand across all games is 0.000. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2: Sensitivity of treatment effects to demand treatments in the effort experiment

	(1) Score	(2) Score (z-scored)
Panel A: Sensitivity		
Positive demand **[1]**	252.342***	0.357***
	(65.307)	(0.092)
Negative demand **[2]**	-344.466***	-0.488***
	(65.030)	(0.092)
1-cent bonus **[3]**	545.549***	0.773***
	(66.799)	(0.095)
Positive demand × 1-cent-bonus **[4]**	-189.547**	-0.268**
	(92.800)	(0.131)
Negative demand × 1-cent-bonus **[5]**	226.574**	0.321**
	(91.188)	(0.129)
Panel B: Bounding		
Conventional treatment effect: **[3]**	545.549***	0.773***
	(66.799)	(0.095)
Lower bound: **[2]** + **[3]** + **[5]** - **[1]**	175.315***	0.248***
	(62.366)	(0.088)
Upper bound: **[1]** + **[3]** + **[4]** - **[2]**	952.811***	1.349***
	(64.136)	(0.091)

Notes: In this table we use data the real effort experiment using strong demand treatments (experiment 3). In Panel A we present the sensitivity of effort to monetary incentives, our demand treatments and interactions of the demand treatments and monetary incentives. In Panel B we display the conventional treatment effects, the lower as well as the upper bound of treatment effects. In column 1 we present the results in terms of raw real-effort, while in column 2 we display z-scored real effort. Robust standard errors in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3: Structural Estimates of Demand Effects

	Power cost of effort			Exponential cost of effort		
	(1) Log Count	(2) Log Count	(3) Log Count	(4) Count	(5) Count	(6) Count
ϕ		0.183**	0.259***		0.213***	0.306***
		(0.091)	(0.091)		(0.078)	(0.064)
$h^L(0)p^L(0)$		-0.720***	-0.488		-0.506***	-0.162
		(0.174)	(0.298)		(0.193)	(0.247)
$h^L(>0)p^L(>0)$		-0.404			0.958	
		(2.072)			(1.732)	
$h^L(1)p^L(1)$			-0.372			0.197
			(1.033)			(0.665)
$h^L(4)p^L(4)$			-6.412**			-6.516***
			(3.082)			(1.864)
s	0.033	0.188**	0.288**	0.031	0.242**	0.529**
	(0.049)	(0.095)	(0.125)	(0.045)	(0.097)	(0.221)
k	6.0e-26	4.2e-23	3.9e-16	4.6e-08	3.0e-06	2.3e-04
	(3.9e-25)	(1.5e-22)	(1.7e-15)	(1.9e-07)	(4.9e-06)	(3.7e-04)
γ	7.228***	6.354***	4.196***	6.5e-03***	4.5e-03***	2.2e-03***
	(2.188)	(1.215)	(1.543)	(2.1e-03)	(8.2e-04)	(7.7e-04)
Observations	729	1699	1699	729	1699	1699
R-squared	0.125	0.167	0.168	0.169	0.205	0.207

Notes: This table uses data from the the real effort experiments on MTurk with strong demand treatments. Coefficients s and ϕ are measured in cents. s measures the respondents intrinsic motivation. ϕ measures the monetary equivalent of acting according to the experimental objective for a worker who is certain about the experimenter's objective. γ is the estimate of the cost curvature (inverse of the elasticity of effort) and k is the scaling parameter. $h^L(0)p^L(0)$ measures latent demand in the no-incentive condition. $h^L(>0)p^L(>0)$ measures latent demand in the 1-cent and 4-cent incentive conditions. $h^L(1)p^L(1)$ measures latent demand in the 1-cent incentive condition. $h^L(4)p^L(4)$ measures latent demand in the 4-cent incentive condition. Robust standard errors in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 4: Response to weak demand treatments across all incentivized games

	Time	Risk	Ambiguity Aversion	Effort 0 cent bonus	Effort 1 cent bonus	Lying	Dictator Game	Ultimatum Game 1	Ultimatum Game 2	Trust Game 1	Trust Game 2
Panel A: Unconditional Means											
Positive demand	0.768	0.524	0.562	0.329	0.484	0.537	0.382	0.473	0.412	0.453	0.398
	(0.027)	(0.023)	(0.024)	(0.014)	(0.012)	(0.012)	(0.014)	(0.014)	(0.013)	(0.023)	(0.017)
No demand		0.541					0.313				
		(0.021)					(0.015)				
Negative demand	0.768	0.469	0.501	0.342	0.468	0.531	0.316	0.443	0.362	0.427	0.346
	(0.026)	(0.020)	(0.024)	(0.014)	(0.013)	(0.011)	(0.014)	(0.013)	(0.013)	(0.025)	(0.012)
Panel B: Sensitivity (positive - negative)											
Raw data	0.000	0.055*	0.060*	-0.013	0.016	0.006	0.066***	0.030	0.050***	0.026	0.051**
	(0.038)	(0.030)	(0.034)	(0.019)	(0.017)	(0.016)	(0.020)	(0.019)	(0.018)	(0.034)	(0.021)
Z-score	0.000	0.164*	0.178*	-0.069	0.084	0.040	0.250***	0.175	0.274***	0.079	0.288**
	(0.096)	(0.091)	(0.101)	(0.101)	(0.094)	(0.102)	(0.075)	(0.113)	(0.101)	(0.104)	(0.120)
		[0.077]					[0.001]				
Panel C: Monotonicity											
Positive - Neutral (z-score)		-0.051					0.260***				
		(0.092)					(0.078)				
		[0.237]					[0.001]				
Negative - Neutral (z-score)		-0.215**					0.010				
		(0.087)					(0.077)				
		[0.041]					[0.426]				
Observations	426	743	393	392	383	413	761	361	413	355	347

Notes: This table uses data from all MTurk experiments with weak demand treatments using real stakes. In Panel A we display the unconditional means and standard errors of those means in the positive, negative and no-demand treatment arms respectively. In Panel B we present the raw and z-scored sensitivity of behavior to our demand treatments. In Panel C we display the sensitivity of behavior in the positive and negative demand condition compared to the no-demand condition. In the demand treatments we reveal the experimental hypothesis to our respondents. Robust standard errors are in parentheses. False-discovery rate adjusted p-values are in brackets. The p-value of an F-test which tests for differences in response to demand across all games is 0.063. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

A Online Appendix: Additional Tables

Table A.1: Moderators of response to strong demand treatments (z-scored)

	All Games	Time	Risk	Ambiguity Aversion	Effort 0 cent bonus	Effort 1 cent bonus	Lying	Dictator Game	Ultimatum Game 1	Ultimatum Game 2	Trust Game 1	Trust Game 2
Panel A: Design Characteristics												
Sensitivity × Incentive	0.058	-0.072	0.183					0.058				
	(0.072)	(0.132)	(0.116)					(0.121)				
Observations	3000	998	1000					1002				
Panel B: Respondent Characteristics												
Sensitivity × Male	-0.143**	-0.222	-0.128	-0.382**	0.040	0.029	-0.213	-0.238	-0.137	-0.156	0.078	-0.344
	(0.057)	(0.167)	(0.125)	(0.192)	(0.196)	(0.211)	(0.217)	(0.151)	(0.188)	(0.200)	(0.216)	(0.239)
Observations	6013	494	1071	404	495	475	366	1118	409	425	383	373
Sensitivity × Attention	0.135	0.311	0.461	-0.276			0.249	-0.043	-0.272	0.228	0.908**	-0.084
	(0.141)	(0.393)	(0.398)	(0.414)			(0.358)	(0.610)	(0.394)	(0.538)	(0.409)	(0.310)
Observations	5043	494	1071	404			366	1118	409	425	383	373
Sensitivity × Representative sample	0.054		-0.127					0.236*				
	(0.098)		(0.130)					(0.141)				
Observations	2189		1071					1118				

Notes: The outcome variables are normalized at the game-interaction term level. In Panel A we display heterogeneous treatment effects of the strong demand treatments by a design characteristics, i.e. whether our respondents' choices are incentivized or hypothetical. In Panel B we display heterogeneous treatment effects by respondent characteristics, namely by gender, attention and whether our respondents come from MTurk or a representative sample. The variable male takes value one if our respondent is male and zero otherwise, attention takes value one if our respondent correctly completed the screener and zero otherwise. The variable representative sample takes value if our respondent come from a representative sample and zero when they come from the MTurk sample. In the strong demand treatments we reveal the experimental objective to our respondents. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.2: Moderators of response to weak demand treatments (z-scored)

	All Games	Time	Risk	Ambiguity Aversion	Effort 0 cent bonus	Effort 1 cent bonus	Lying	Dictator Game	Ultimatum Game 1	Ultimatum Game 2	Trust Game 1	Trust Game 2
Panel A: Design Characteristics												
Sensitivity × Incentive	0.079		0.140					0.017				
	(0.085)		(0.125)					(0.115)				
Observations	1976		978					998				
Panel B: Respondent Characteristics												
Sensitivity × Male	-0.001	-0.020	-0.070	0.059	0.349	0.035	0.064	-0.137	-0.113	0.015	0.282	-0.218
	(0.058)	(0.173)	(0.133)	(0.203)	(0.238)	(0.232)	(0.188)	(0.148)	(0.193)	(0.185)	(0.229)	(0.229)
Observations	5618	426	1046	393	392	383	413	1089	361	413	355	347
Sensitivity × Attention	0.147	-0.413	-0.067	0.440			0.224	0.692*	0.105	0.364	0.119	-0.389
	(0.124)	(0.395)	(0.305)	(0.504)			(0.305)	(0.377)	(0.296)	(0.230)	(0.362)	(0.301)
Observations	4843	426	1046	393			413	1089	361	413	355	347
Sensitivity × Representative sample	0.026		0.028					0.026				
	(0.100)		(0.140)					(0.139)				
Observations	2135		1046					1089				

Notes: The outcome variables are normalized at the game-interaction term level. In Panel A we display heterogeneous treatment effects of the strong demand treatments by a design characteristics, i.e. whether our respondents' choices are incentivized or hypothetical. In Panel B we display heterogeneous treatment effects by respondent characteristics, namely by gender, attention and whether our respondents come from MTurk or a representative sample. The variable male takes value one if our respondent is male and zero otherwise, attention takes value one if our respondent correctly completed the screener and zero otherwise. The variable representative sample takes value if our respondent come from a representative sample and zero when they come from the MTurk sample. In the weak demand treatments we reveal the experimental hypothesis to our respondents. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.3: Overview of Experiments

Experiment	**Sample**	**Games**	**Demand Treatments**	**Real or Hypothetical**
Experiment 1	MTurk (N=4495)	Dictator Game, Investment Game and Convex Time Budgets	Strong positive demand, strong negative demand and no-demand treatment	Both real stakes and hypothetical choices
Experiment 2	MTurk (N=2964)	Dictator Game and Investment Game	Weak positive demand, weak negative demand and no-demand treatment	Both real stakes and hypothetical choices
Experiment 3	MTurk (N=1452)	Effort experiment with 1 cent bonus and Effort experiment with no bonus	Strong positive, strong negative and no-demand treatment	Real stakes (real effort experiment)
Experiment 4	Representative online Panel (N=2941)	Dictator Game and Investment Game	Strong positive demand, strong negative demand, weak positive demand and weak negative demand and no-demand treatment	Real stakes
Experiment 5	MTurk (N=5068)	Trust game (first and second mover), Ultimatum game (first and second mover), Coinflip game, Investment Game with uncertainty (ambiguity aversion) and Convex Time Budgets	Strong positive demand, strong negative demand, weak positive demand and weak negative demand	Real stakes
Experiment 6	MTurk (N=775)	Effort experiment with 1 cent bonus, Effort experiment with no bonus	Weak positive demand and weak negative demand	Real stakes (real effort experiment)

Notes: This table summarizes the key design features of each of the experiments. In experiment 5 for convex time budgets we only employ the weak demand treatments, while for all other games we employ both strong and weak demand treatments. In experiment 2 we also have an additional incentive treatment arm in which respondents receive four cents per 100 tasks in which we do not induce any additional demand.

Table A.4: Design by Games: Strong Demand Experiments

Game	Description	Sample	Show-up fee	Choice set	Demand Instructions
Dictator Game	Choose to split money between yourself and another participant.	MTurk (N=1,508) Representative Sample (N=899)	$.25	action ∈ [0,1]	"You will do us a favor if you give more (less) to the other participant than you normally would."
Investment Game	Choose to how much to invest in a risky project.	MTurk (N=1,499) Representative Sample (N=902)	$.25	action ∈ [0,1]	"You will do us a favor if you invest more (less) than you normally would."
Investment Game with ambiguous returns	Choose to how much to invest in a project with uncertain returns.	MTurk (N= 404)	$.25	action ∈ [0,1]	"You will do us a favor if you invest more (less) than you normally would."
Convex Time Budgets	Choose between receiving money today vs. money in seven days.	MTurk (N=1,488); Rep. Sample (N=899)	$.25	action ∈ [0,1.2]	"You will do us a favor if you choose to receive more (less) in seven days than you normally would."
Effort: No bonus	Alternately press the a and b button without receiving any bonus.	MTurk (N=735)	$1	action ∈ [0,4000]	"You will do us a favor if you work harder (less hard) than you normally would."
Effort: 1-cent bonus	Alternately press the a and b button while receiving 1 cent per 100 scores.	MTurk (N=717)	$1	action ∈ [0,4000]	"You will do us a favor if you work harder (less hard) than you normally would."
Trust Game 1st mover	Choose to send an amount of money to the other player.	MTurk (N=383)	$.25	action ∈ [0,.2,.4,.6,.8,1]	"You will do us a favor if you send more (less) to the other participant than you normally would."
Trust Game 2nd mover	Choose to send back some money to the other player. (Strategy method)	MTurk (N=373)	$.25	action ∈ [0,1.2]	"You will do us a favor if you send back more (less) to the other participant than you normally would."
Ultimatum Game 1st mover	Offer a split to the other player.	MTurk (N=409)	$.25	action ∈ [0,1]	"You will do us a favor if you offer more (less) to the other participant than you normally would."
Ultimatum Game 2nd mover	Choose to send back some money to the other player.	MTurk (N=425)	$.25	action ∈ [0,1]	"You will do us a favor if you require a higher (lower) minimum amount than you normally would."
Lying	Report the number of "Heads" after 10 coin-flips.	MTurk (N=366)	$.25	action ∈ [0,1,2,3,4,5,6,7,8,9,10]	"You will do us a favor if you report more (fewer) heads than you normally would."

Notes: This table summarizes the key design features of each of the games.

Table A.5: Design by Games: Weak Demand Experiments

Game	**Description**	**Sample**	**Show-up fee**	**Choice set**	**Demand Instructions**
Dictator Game	Choose to split money between yourself and another participant.	MTurk (N=1,482) Representative Sample (N=865)	$.25	action ∈ [0,1]	"We expect that participants who are shown these instructions will give more (less) to the other participant than they normally would."
Investment Game	Choose to how much to invest in a risky project.	MTurk (N=1,482) Representative Sample (N=883)	$.25	action ∈ [0,1]	"We expect that participants who are shown these instructions will invest more (less) than they normally would."
Investment Game with ambiguous returns	Choose to how much to invest in a project with uncertain returns.	MTurk (N= 393)	$.25	action ∈ [0,1]	"We expect that participants who are shown these instructions will invest more (less) than they normally would."
Convex Time Budgets	Choose between receiving money today vs. money in seven days.	MTurk (N=426)	$.25	action ∈ [0,1.2]	"We expect that participants who are shown these instructions will choose to receive more (less) in seven days than they normally would."
Effort: No bonus	Alternately press the a and b button without receiving any bonus.	MTurk (N=392)	$1	action ∈ [0,4000]	"We expect that participants who are shown these instructions will work harder (less hard) than they normally would."
Effort: 1-cent bonus	Alternately press the a and b button while receiving 1 cent per 100 scores.	MTurk (N=383)	$1	action ∈ [0,4000]	"We expect that participants who are shown these instructions will work harder (less hard) than they normally would."
Trust Game 1st mover	Choose to send an amount of money to the other player.	MTurk (N=355)	$.25	action ∈ [0,.2,.4,.6,.8,1]	"We expect that participants who are shown these instructions will send more (less) to the other participant than they normally would."
Trust Game 2nd mover	Choose to send back some money to the other player. (Strategy method)	MTurk (N=347)	$.25	action ∈ [0,1.2]	"We expect that participants who are shown these instructions will send back more (less) to the other participant than they normally would."
Ultimatum Game 1st mover	Offer a split to the other player.	MTurk (N=361)	$.25	action ∈ [0,1]	"We expect that participants who are shown these instructions will offer more (less) to the other participant than they normally would."
Ultimatum Game 2nd mover	Choose to send back some money to the other player.	MTurk (N=413)	$.25	action ∈ [0,1]	"We expect that participants who are shown these instructions will require a higher (lower) minimum amount than they normally would."
Lying	Report the number of "Heads" after 10 coin-flips.	MTurk (N=413)	$.25	action ∈ [0,1,2,3,4,5,6,7,8,9,10]	"We expect that participants who are shown these instructions will report more (fewer) heads than they normally would."

Notes: This table summarizes the key design features of each of the games.

Table A.6: Predicted Values from Structural Model

	Power effort cost			Exponential effort cost		
	(1)	(2)	(3)	(4)	(5)	(6)
	$\log(a^L(\zeta))$	$\log(a(\zeta))$	$\log(a(\zeta))$	$a^L(\zeta)$	$a(\zeta)$	$a(\zeta)$
0 cents	6.93	7.12	7.07	1363	1495	1407
1 cent	7.40	7.41	7.42	1904	1860	1886
4 cents	7.59	7.61	7.71	2114	2134	2376

Columns 1–3 present predicted values from the power effort cost model, and 4–6 for the exponential cost model. Column numbers correspond to those in table 3. Rows correspond to incentive treatments, in cents per 100 points. Therefore (1) and (4) are predicted values from the model without demand effects, equalling mean observed actions under the neutral treatments, and are potentially contaminated by latent demand. Columns (2) and (5) are predicted demand-free actions when latent demand is restricted to be equal in the 1 and 4 cent treatments. Columns (3) and (6) are predicted demand-free actions when latent demand is allowed to differ across all treatments.

Table A.7: Confidence interval for the interval and the parameter

	Time	Risk	Ambiguity Aversion	Effort 0 cent bonus	Effort 1 cent bonus	Lying	Dictator Game	Ultimatum Game 1	Ultimatum Game 2	Trust Game 1	Trust Game 2
Panel A: Strong Demand											
Interval	[0.659, 0.792]	[0.373, 0.548]	[0.428, 0.583]	[0.254, 0.403]	[0.447, 0.492]	[0.447, 0.492]	[0.252, 0.433]	[0.404, 0.520]	[0.338, 0.474]	[0.350, 0.532]	[0.288, 0.470]
95% CI on interval	[0.612, 0.831]	[0.342, 0.581]	[0.391, 0.622]	[0.235, 0.422]	[0.429, 0.511]	[0.487, 0.625]	[0.228, 0.458]	[0.381, 0.541]	[0.314, 0.496]	[0.314, 0.571]	[0.263, 0.499]
95% CI on parameter	[0.622, 0.823]	[0.349, 0.574]	[0.399, 0.613]	[0.240, 0.418]	[0.433, 0.507]	[0.493, 0.621]	[0.233, 0.452]	[0.386, 0.536]	[0.319, 0.491]	[0.322, 0.562]	[0.269, 0.493]
Observations	730	730	404	735	717	366	773	409	425	383	373
Panel B: Weak Demand											
Interval	[0.768, 0.768]	[0.469, 0.524]	[0.501, 0.562]	[0.342, 0.329]	[0.468, 0.484]	[0.468, 0.484]	[0.316, 0.382]	[0.443, 0.473]	[0.362, 0.412]	[0.427, 0.453]	[0.346, 0.398]
95% CI on interval	[0.716, 0.820]	[0.436, 0.561]	[0.462, 0.601]	[0.315, 0.356]	[0.447, 0.504]	[0.511, 0.557]	[0.293, 0.405]	[0.422, 0.496]	[0.342, 0.435]	[0.385, 0.492]	[0.326, 0.426]
95% CI on parameter	[0.724, 0.812]	[0.443, 0.553]	[0.471, 0.593]	[0.320, 0.352]	[0.452, 0.500]	[0.514, 0.553]	[0.298, 0.400]	[0.427, 0.491]	[0.346, 0.430]	[0.393, 0.484]	[0.330, 0.420]
Observations	426	743	393	392	383	413	761	361	413	355	347

Notes: This table uses data from all MTurk experiments with strong and weak demand treatments using real stakes. This table shows the 95 percent confidence interval for the parameter and the interval respectively. We provide a Stata package, `demandbounds`, which computes the upper and lower bound of confidence intervals for mean behavior and treatment effects using the method proposed by Imbens and Manski (2004). This bounding exercise is based on strong and weak demand treatments in which we manipulate our participants' belief about the experimental objective and hypothesis respectively.

Table A.8: Confidence intervals for treatment effects

	Treatment Effect: Score in Effort Task
Interval	[175.315, 952.811]
95% CI on interval	[72.733, 1058.305]
95% CI on parameter	[95.390, 1035.004]
Observations	1452

Notes: In this table we use data from the real effort experiment using strong demand treatments (experiment 3). This table shows the 95 percent confidence interval for the parameter and the set respectively. Our estimates are based on a Stata package, `demandbounds`, which computes the upper and lower bound of confidence intervals for mean behavior and treatment effects using the method proposed by Imbens and Manski (2004). This bounding exercise is based on strong demand treatments in which we manipulate our participants' belief about the experimental objective.

Table A.9: Heterogeneous Reponse to the strong demand treatment (raw choices)

	All Games	Time	Risk	Ambiguity Aversion	Effort 0 cent bonus	Effort 1 cent bonus	Lying	Dictator Game	Ultimatum Game 1	Ultimatum Game 2	Trust Game 1	Trust Game 2
Panel A: Design Characteristics												
Sensitivity × Incentive	0.026	-0.028	0.062					0.014				
	(0.029)	(0.051)	(0.039)					(0.030)				
Observations	3000	998	1000					1002				
Panel B: Respondent Characteristics												
Sensitivity × Male	-0.047***	-0.098	-0.041	-0.130**	0.007	0.005	-0.037	-0.056	-0.028	-0.031	0.023	-0.066
	(0.016)	(0.074)	(0.040)	(0.065)	(0.032)	(0.034)	(0.037)	(0.035)	(0.038)	(0.040)	(0.065)	(0.045)
Observations	6013	494	1071	404	495	475	366	1118	409	425	383	373
Sensitivity × Attention	0.019	0.131	0.150***	-0.080			0.059	-0.001	-0.079	0.071	0.270**	-0.019
	(0.026)	(0.166)	(0.042)	(0.119)			(0.084)	(0.036)	(0.114)	(0.168)	(0.122)	(0.070)
Observations	5043	494	1071	404			366	1118	409	425	383	373
Sensitivity × Representative sample	0.015		-0.038					0.053*				
	(0.025)		(0.039)					(0.032)				
Observations	2189		1071					1118				

Notes: In Panel A we display heterogeneous treatment effects of the strong demand treatments by a design characteristics, i.e. whether our respondents' choices are incentivized or hypothetical. In Panel B we display heterogeneous treatment effects by respondent characteristics, namely by gender, attention and whether our respondents come from MTurk or a representative sample. The variable male takes value one if our respondent is male and zero otherwise, attention takes value one if our respondent correctly completed the screener and zero otherwise. The variable representative sample takes value if our respondent come from a representative sample and zero when they come from the MTurk sample. In the strong demand treatments we reveal the experimental objective to our respondents. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.10: Heterogeneous Reponse to the weak demand treatment (raw choices)

	All Games	Time	Risk	Ambiguity Aversion	Effort 0 cent bonus	Effort 1 cent bonus	Lying	Dictator Game	Ultimatum Game 1	Ultimatum Game 2	Trust Game 1	Trust Game 2
Panel A: Design Characteristics												
Sensitivity × Incentive	0.026		0.048					0.004				
	(0.027)		(0.042)					(0.028)				
Observations	1976		978					998				
Panel B: Respondent Characteristics												
Sensitivity × Male	0.000	-0.009	-0.022	0.020	0.057	0.006	0.011	-0.031	-0.023	0.003	0.085	-0.042
	(0.016)	(0.077)	(0.042)	(0.069)	(0.039)	(0.037)	(0.033)	(0.034)	(0.039)	(0.037)	(0.069)	(0.044)
Observations	5618	426	1046	393	392	383	413	1089	361	413	355	347
Sensitivity × Attention	0.007	-0.174	0.018	0.127			0.053	0.047	0.030	0.114	0.035	-0.087
	(0.022)	(0.167)	(0.043)	(0.145)			(0.072)	(0.035)	(0.086)	(0.072)	(0.108)	(0.068)
Observations	4843	426	1046	393			413	1089	361	413	355	347
Sensitivity × Representative sample	0.007		0.009					0.006				
	(0.027)		(0.042)					(0.031)				
Observations	2135		1046					1089				

Notes: In Panel A we display heterogeneous treatment effects of the strong demand treatments by a design characteristics, i.e. whether our respondents' choices are incentivized or hypothetical. In Panel B we display heterogeneous treatment effects by respondent characteristics, namely by gender, attention and whether our respondents come from MTurk or a representative sample. The variable male takes value one if our respondent is male and zero otherwise, attention takes value one if our respondent correctly completed the screener and zero otherwise. The variable representative sample takes value if our respondent come from a representative sample and zero when they come from the MTurk sample. In the strong demand treatments we reveal the experimental objective to our respondents. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.11: Additional Heterogeneity: Strong demand treatment (z-scored)

	All Games	Time	Risk	Ambiguity Aversion	Effort 0 cent bonus	Effort 1 cent bonus	Lying	Dictator Game	Ultimatum Game 1	Ultimatum Game 2	Trust Game 1	Trust Game 2
Sensitivity × High Education	0.015	-0.153	0.163	-0.213	0.100	-0.334*	0.397	-0.057	-0.153	0.132	0.079	0.113
	(0.073)	(0.193)	(0.173)	(0.195)	(0.171)	(0.187)	(0.247)	(0.168)	(0.226)	(0.221)	(0.206)	(0.266)
Observations	6330	998	1000	404	495	475	366	1002	409	425	383	373
Sensitivity × Experienced	0.114	0.087	0.018	0.049			0.196	-0.139	0.128	0.105	0.160	0.032
	(0.084)	(0.194)	(0.177)	(0.206)			(0.242)	(0.164)	(0.227)	(0.237)	(0.207)	(0.257)
Observations	5043	494	1071	404			366	1118	409	425	383	373

Notes: Our outcome measures are normalized at the game level using the negative demand condition. We display heterogeneous treatment effects by respondent characteristics, namely by education and experience. High Education takes value one if a respondent has at least a bachelor degree. Experienced takes value one if a respondent has completed at least 4000 HITs on MTurk. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.12: Additional Heterogeneity: Weak demand treatment (z-scored)

	All Games	Time	Risk	Ambiguity Aversion	Effort 0 cent bonus	Effort 1 cent bonus	Lying	Dictator Game	Ultimatum Game 1	Ultimatum Game 2	Trust Game 1	Trust Game 2
Sensitivity × High Education	-0.064	0.037	0.055	-0.009	-0.144	0.060	-0.137	0.290*	-0.319	0.322	-0.191	-0.283
	(0.068)	(0.192)	(0.185)	(0.210)	(0.210)	(0.201)	(0.213)	(0.149)	(0.221)	(0.204)	(0.218)	(0.243)
Observations	5459	426	978	393	392	383	413	998	361	413	355	347
Sensitivity × Experienced	-0.059	0.148	0.100	-0.142			-0.051	-0.214	-0.203	-0.304	-0.005	0.109
	(0.080)	(0.205)	(0.226)	(0.221)			(0.205)	(0.184)	(0.231)	(0.208)	(0.232)	(0.250)
Observations	4843	426	1046	393			413	1089	361	413	355	347

Notes: Our outcome measures are normalized at the game level using the negative demand condition. We display heterogeneous treatment effects by respondent characteristics, namely by education and experience. High Education takes value one if a respondent has at least a bachelor degree. Experienced takes value one if a respondent has completed at least 4000 HITs on MTurk. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.13: Belief about the experimental objective in response to the strong demand treatments

	Belief: Time	Belief: Risk	Belief: Ambiguity Aversion	Belief: Effort 0 cent bonus	Belief: Effort 1 cent bonus	Belief: Lying	Belief: Dictator Game	Belief: Ult. Game 1	Belief: Ult. Game 2	Belief: Trust Game 1	Belief: Trust Game 2
Panel A: Unconditional Means											
Positive demand	0.797	0.702	0.701	0.773	0.942	0.811	0.648	0.572	0.662	0.410	0.386
	(0.025)	(0.030)	(0.032)	(0.027)	(0.015)	(0.029)	(0.029)	(0.034)	(0.032)	(0.035)	(0.036)
No demand	0.720	0.534		0.733	0.888		0.354				
	(0.029)	(0.032)		(0.020)	(0.020)		(0.030)				
Negative demand	0.622	0.424	0.335	0.294	0.509	0.562	0.243	0.309	0.359	0.295	0.217
	(0.031)	(0.031)	(0.033)	(0.029)	(0.033)	(0.037)	(0.028)	(0.033)	(0.033)	(0.034)	(0.030)
Panel B: Sensitivity (Positive - Negative)											
Raw data	0.175***	0.278***	0.366***	0.479***	0.434***	0.248***	0.405***	0.263***	0.303***	0.115**	0.169***
	(0.040)	(0.043)	(0.046)	(0.039)	(0.036)	(0.047)	(0.040)	(0.047)	(0.046)	(0.049)	(0.047)
Z-score	0.360***	0.557***	0.773***	1.051***	0.866***	0.499***	0.899***	0.567***	0.632***	0.251**	0.409***
	(0.083)	(0.087)	(0.098)	(0.086)	(0.072)	(0.095)	(0.089)	(0.102)	(0.097)	(0.106)	(0.113)
	[0.001]	[0.001]		[0.001]	[0.001]		[0.001]				
Panel C: Monotonicity											
Positive - Neutral (z-score)	0.157**	0.335***		0.088	0.108**		0.653***				
	(0.079)	(0.088)		(0.073)	(0.050)		(0.092)				
	[0.033]	[0.001]		[0.082]	[0.011]		[0.001]				
Negative - Neutral (z-score)	-0.202**	-0.222**		-0.962***	-0.758***		-0.246***				
	(0.088)	(0.089)		(0.077)	(0.077)		(0.090)				
	[0.022]	[0.004]		[0.001]	[0.001]		[0.002]				
Observations	730	730	404	982	717	366	773	409	425	383	373

Notes: The outcome variables take value one of the respondents believed that the experimenter wanted a high action. In Panel A we display the unconditional means and standard errors of those means in the positive, negative and no-demand treatment arms respectively. In Panel B we present the raw and z-scored sensitivity of beliefs to our demand treatments. In Panel C we display the sensitivity of beliefs in the positive and negative demand condition compared to the no-demand condition. In the demand treatments we reveal the experimental objective to our respondents. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.14: Belief about the experimental objective in response to the weak demand treatments

	Belief: Time	Belief: Risk	Belief: Ambiguity Aversion	Belief: Effort 0 cent bonus	Belief: Effort 1 cent bonus	Belief: Lying	Belief: Dictator Game	Belief: Ult. Game 1	Belief: Ult. Game 2	Belief: Trust Game 1	Belief: Trust Game 2
Panel A: Unconditional Means											
Positive demand	0.829	0.757	0.763	0.790	0.979	0.779	0.540	0.700	0.684	0.611	0.669
	(0.026)	(0.028)	(0.030)	(0.029)	(0.010)	(0.028)	(0.032)	(0.034)	(0.032)	(0.036)	(0.038)
No demand		0.620					0.321				
		(0.030)					(0.030)				
Negative demand	0.602	0.372	0.328	0.284	0.356	0.464	0.231	0.238	0.382	0.112	0.083
	(0.033)	(0.031)	(0.034)	(0.032)	(0.035)	(0.036)	(0.026)	(0.032)	(0.034)	(0.024)	(0.020)
Panel B: Sensitivity (Positive - Negative)											
Raw data	0.227***	0.386***	0.434***	0.505***	0.623***	0.315***	0.309***	0.462***	0.303***	0.499***	0.586***
	(0.042)	(0.042)	(0.045)	(0.044)	(0.036)	(0.046)	(0.041)	(0.047)	(0.047)	(0.043)	(0.043)
Z-score	0.466***	0.772***	0.918***	1.109***	1.244***	0.633***	0.685***	0.998***	0.631***	1.091***	1.418***
	(0.087)	(0.084)	(0.096)	(0.095)	(0.072)	(0.092)	(0.090)	(0.101)	(0.098)	(0.095)	(0.104)
		[0.001]					[0.001]				
Panel C: Monotonicity											
Positive - Neutral (z-score)		0.274***					0.486***				
		(0.082)					(0.097)				
		[0.001]					[0.001]				
Negative - Neutral (z-score)		-0.497***					-0.199**				
		(0.086)					(0.088)				
		[0.001]					[0.008]				
Observations	426	743	393	392	383	413	761	361	413	355	347

Notes: The outcome variables take value one of the respondents believed that the experimenter wanted a high action. In Panel A we display the unconditional means and standard errors of those means in the positive, negative and no-demand treatment arms respectively. In Panel B we present the raw and z-scored sensitivity of beliefs to our demand treatments. In Panel C we display the sensitivity of beliefs in the positive and negative demand condition compared to the no-demand condition. In the demand treatments we reveal the experimental hypothesis to our respondents. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.15: Belief about the experimental hypothesis in response to the strong demand treatments

	Belief: Time	Belief: Risk	Belief: Ambiguity Aversion	Belief: Effort 0 cent bonus	Belief: Effort 1 cent bonus	Belief: Lying	Belief: Dictator Game	Belief: Ult. Game 1	Belief: Ult. Game 2	Belief: Trust Game 1	Belief: Trust Game 2
Panel A: Unconditional Means											
Positive demand	0.727	0.592	0.593	0.729	0.934	0.789	0.452	0.670	0.653	0.580	0.587
	(0.028)	(0.033)	(0.034)	(0.028)	(0.016)	(0.030)	(0.030)	(0.032)	(0.032)	(0.035)	(0.036)
No demand	0.682	0.482		0.700	0.855		0.142				
	(0.030)	(0.032)		(0.021)	(0.023)		(0.022)				
Negative demand	0.639	0.420	0.400	0.266	0.573	0.625	0.185	0.284	0.440	0.290	0.243
	(0.031)	(0.031)	(0.035)	(0.028)	(0.033)	(0.037)	(0.025)	(0.032)	(0.034)	(0.034)	(0.031)
Panel B: Sensitivity (Positive - Negative)											
Raw data	0.088**	0.172***	0.193***	0.463***	0.361***	0.164***	0.267***	0.386***	0.213***	0.290***	0.344***
	(0.042)	(0.045)	(0.049)	(0.040)	(0.036)	(0.047)	(0.039)	(0.046)	(0.047)	(0.049)	(0.048)
Z-score	0.181**	0.345***	0.393***	1.051***	0.724***	0.339***	0.624***	0.859***	0.428***	0.638***	0.799***
	(0.086)	(0.090)	(0.100)	(0.091)	(0.073)	(0.097)	(0.092)	(0.102)	(0.095)	(0.107)	(0.112)
	[0.122]	[0.001]		[0.001]	[0.001]		[0.001]				
Panel C: Monotonicity											
Positive - Neutral (z-score)	0.091	0.221**		0.065	0.158***		0.724***				
	(0.085)	(0.091)		(0.080)	(0.056)		(0.087)				
	[0.268]	[0.016]		[0.161]	[0.001]		[0.001]				
Negative - Neutral (z-score)	-0.090	-0.124		-0.987***	-0.566***		0.100				
	(0.090)	(0.089)		(0.079)	(0.080)		(0.077)				
	[0.268]	[0.057]		[0.001]	[0.001]		[0.069]				
Observations	730	730	404	982	717	366	773	409	425	383	373

Notes: The outcome variables take value one of the respondents believed that the experimenter expected a high action. In Panel A we display the unconditional means and standard errors of those means in the positive, negative and no-demand treatment arms respectively. In Panel B we present the raw and z-scored sensitivity of beliefs to our demand treatments. In Panel C we display the sensitivity of beliefs in the positive and negative demand condition compared to the no-demand condition. In the demand treatments we reveal the experimental objective to our respondents. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.16: Belief about the experimental hypothesis in response to the weak demand treatments

	Belief: Time	Belief: Risk	Belief: Ambiguity Aversion	Belief: Effort 0 cent bonus	Belief: Effort 1 cent bonus	Belief: Lying	Belief: Dictator Game	Belief: Ult. Game 1	Belief: Ult. Game 2	Belief: Trust Game 1	Belief: Trust Game 2
Panel A: Unconditional Means											
Positive demand	0.790	0.749	0.707	0.779	0.964	0.871	0.464	0.728	0.733	0.627	0.682
	(0.028)	(0.028)	(0.032)	(0.030)	(0.014)	(0.023)	(0.032)	(0.033)	(0.031)	(0.036)	(0.038)
No demand		0.534					0.160				
		(0.031)					(0.024)				
Negative demand	0.454	0.260	0.221	0.239	0.325	0.526	0.104	0.354	0.362	0.294	0.181
	(0.034)	(0.028)	(0.030)	(0.030)	(0.034)	(0.036)	(0.019)	(0.036)	(0.033)	(0.035)	(0.028)
Panel B: Sensitivity (Positive - Negative)											
Raw data	0.337***	0.489***	0.487***	0.541***	0.639***	0.345***	0.360***	0.374***	0.371***	0.333***	0.500***
	(0.044)	(0.040)	(0.044)	(0.043)	(0.037)	(0.042)	(0.037)	(0.049)	(0.046)	(0.050)	(0.047)
Z-score	0.693***	0.978***	0.991***	1.229***	1.282***	0.712***	0.841***	0.832***	0.746***	0.732***	1.163***
	(0.091)	(0.080)	(0.090)	(0.097)	(0.073)	(0.087)	(0.086)	(0.108)	(0.092)	(0.110)	(0.109)
		[0.001]					[0.001]				
Panel C: Monotonicity											
Positive - Neutral (z-score)		0.431***					0.710***				
		(0.084)					(0.092)				
		[0.001]					[0.001]				
Negative - Neutral (z-score)		-0.548***					-0.131*				
		(0.083)					(0.070)				
		[0.001]					[0.021]				
Observations	426	743	393	392	383	413	761	361	413	355	347

Notes: The outcome variables take value one of the respondents believed that the experimenter expected a high action. In Panel A we display the unconditional means and standard errors of those means in the positive, negative and no-demand treatment arms respectively. In Panel B we present the raw and z-scored sensitivity of beliefs to our demand treatments. In Panel C we display the sensitivity of beliefs in the positive and negative demand condition compared to the no-demand condition. In the demand treatments we reveal the experimental hypothesis to our respondents. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.17: Attrition overview by game in the strong demand experiments

	Finished: Time	Finished: Risk	Finished: Ambiguity Aversion	Finished: Effort 0 cent bonus	Finished: Effort 1 cent bonus	Finished: Lying	Finished: Dictator Game	Finished: Ult. Game 1	Finished: Ult. Game 2	Finished: Trust Game 1	Finished: Trust Game 2
Panel A: Unconditional Means											
Positive demand	0.996	0.996	0.995	0.969	0.968	0.995	1.000	1.000	0.991	0.990	1.000
	(0.004)	(0.004)	(0.005)	(0.011)	(0.011)	(0.005)	(0.000)	(0.000)	(0.006)	(0.007)	(0.000)
No demand	0.996	1.000		0.938	0.980		0.996				
	(0.004)	(0.000)		(0.015)	(0.009)		(0.004)				
Negative demand	0.992	0.992	1.000	0.980	0.963	0.994	0.992	0.985	0.991	1.000	1.000
	(0.006)	(0.005)	(0.000)	(0.009)	(0.012)	(0.006)	(0.006)	(0.009)	(0.007)	(0.000)	(0.000)
Panel B: Differential attrition											
Positive - Negative	0.004	0.003	-0.005	-0.012	0.005	0.000	0.008	0.015*	0.000	-0.010	0.000
	(0.007)	(0.007)	(0.005)	(0.014)	(0.017)	(0.008)	(0.006)	(0.009)	(0.009)	(0.007)	(0.000)
Positive - Neutral	0.000	-0.004		0.031*	-0.012		0.004				
	(0.006)	(0.004)		(0.019)	(0.014)		(0.004)				
Negative - Neutral	-0.004	-0.008		0.043**	-0.017		-0.004				
	(0.007)	(0.005)		(0.018)	(0.015)		(0.007)				
Observations	734	733	405	764	739	368	776	412	429	385	373

Notes: In Panel A we present unconditional the proportion of respondents who completed the experiment in the positive, negative and no-demand treatment arms respectively. In Panel B we assess whether there was differential attrition across treatment arms by examining differences in completion rates across demand treatment arms. In the demand treatments we reveal the experimental objective to our respondents. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

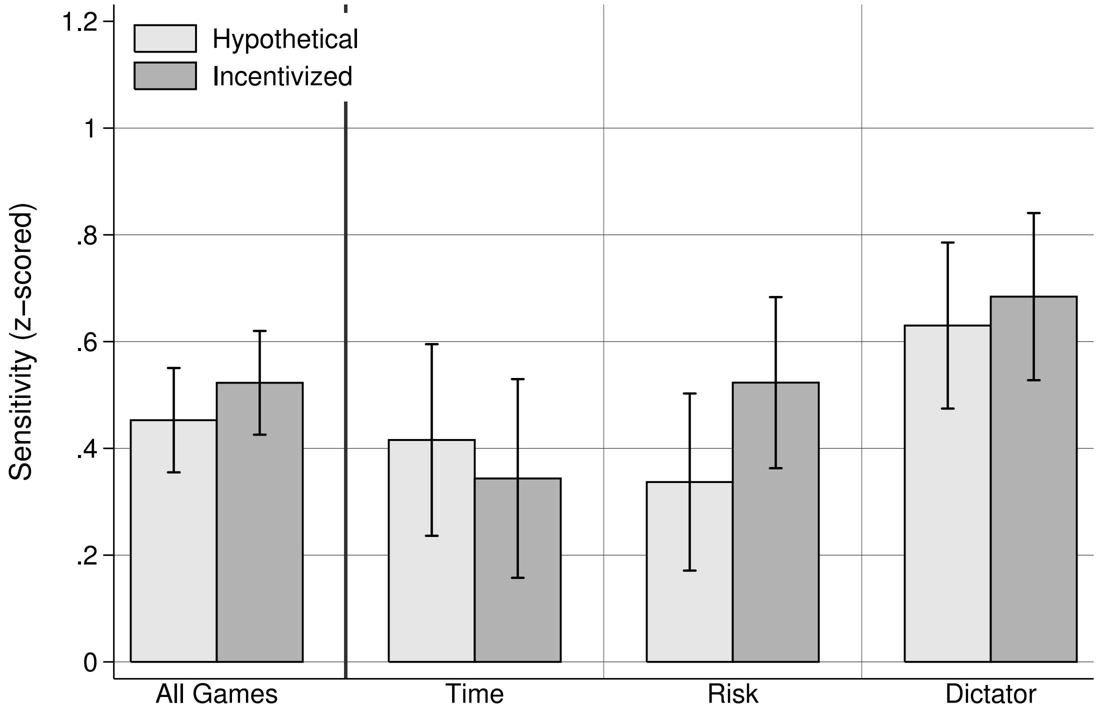
Table A.18: Attrition overview by game in the weak demand experiments

	Finished: Time	Finished: Risk	Finished: Ambiguity Aversion	Finished: Effort 0 cent bonus	Finished: Effort 1 cent bonus	Finished: Lying	Finished: Dictator Game	Finished: Ult. Game 1	Finished: Ult. Game 2	Finished: Trust Game 1	Finished: Trust Game 2
Panel A: Unconditional Means											
Positive demand	0.991	0.987	0.985	0.951	0.937	0.991	0.992	0.994	1.000	1.000	1.000
	(0.007)	(0.007)	(0.009)	(0.015)	(0.017)	(0.006)	(0.006)	(0.006)	(0.000)	(0.000)	(0.000)
No demand		0.993					0.980				
		(0.005)					(0.009)				
Negative demand	0.986	0.984	0.990	0.966	0.955	0.995	0.989	0.989	0.986	1.000	0.980
	(0.008)	(0.008)	(0.007)	(0.013)	(0.015)	(0.005)	(0.006)	(0.008)	(0.008)	(0.000)	(0.010)
Panel B: Differential attrition											
Positive - Negative	0.004	0.004	-0.005	-0.014	-0.018	-0.004	0.003	0.005	0.014*	0.000	0.020**
	(0.010)	(0.011)	(0.011)	(0.020)	(0.023)	(0.008)	(0.008)	(0.009)	(0.008)	(0.000)	(0.010)
Positive - Neutral		-0.005					0.012				
		(0.009)					(0.011)				
Negative - Neutral		-0.009					0.009				
		(0.010)					(0.011)				
Observations	431	752	398	409	405	416	771	364	416	355	351

Notes: In Panel A we present unconditional the proportion of respondents who completed the experiment in the positive, negative and no-demand treatment arms respectively. In Panel B we assess whether there was differential attrition across treatment arms by examining differences in completion rates across demand treatment arms. In the demand treatments we reveal the experimental objective to our respondents. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

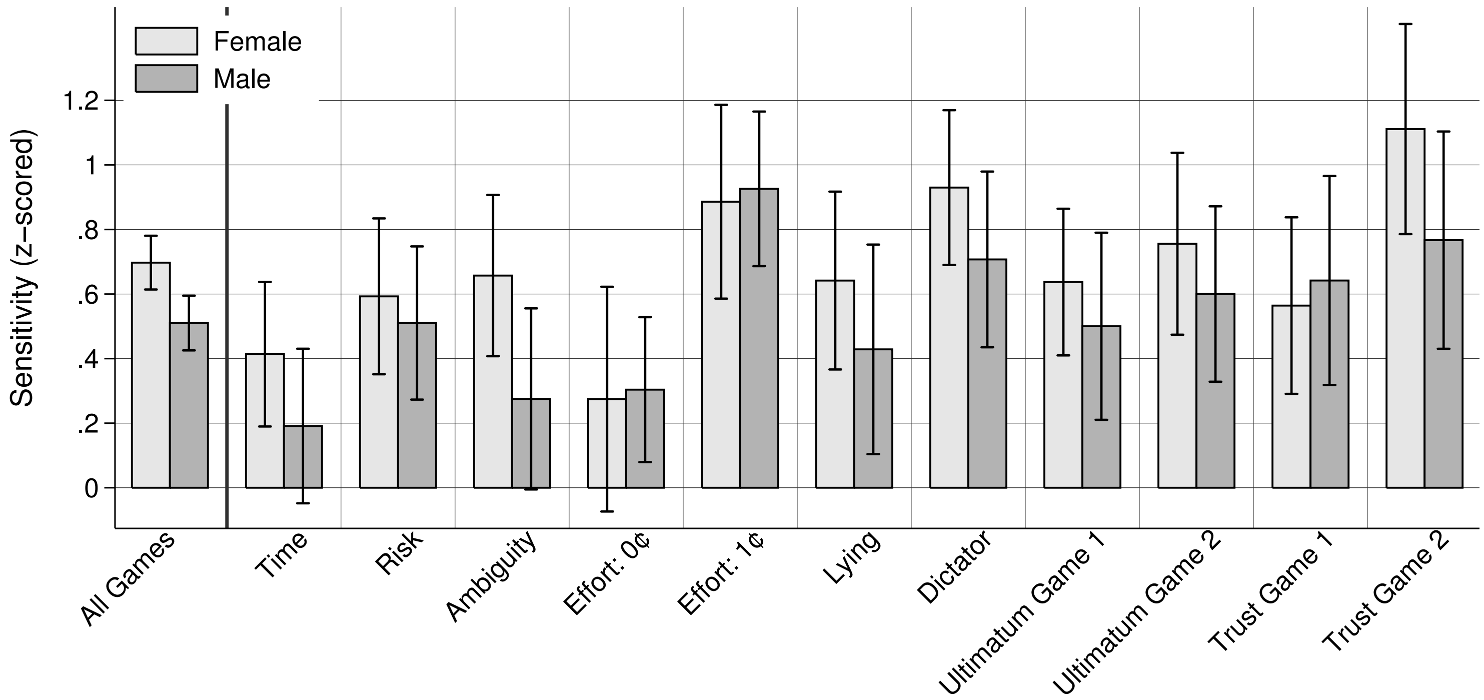
B Online Appendix: Additional Figures

Figure A.1: Response to strong demand treatments by incentives



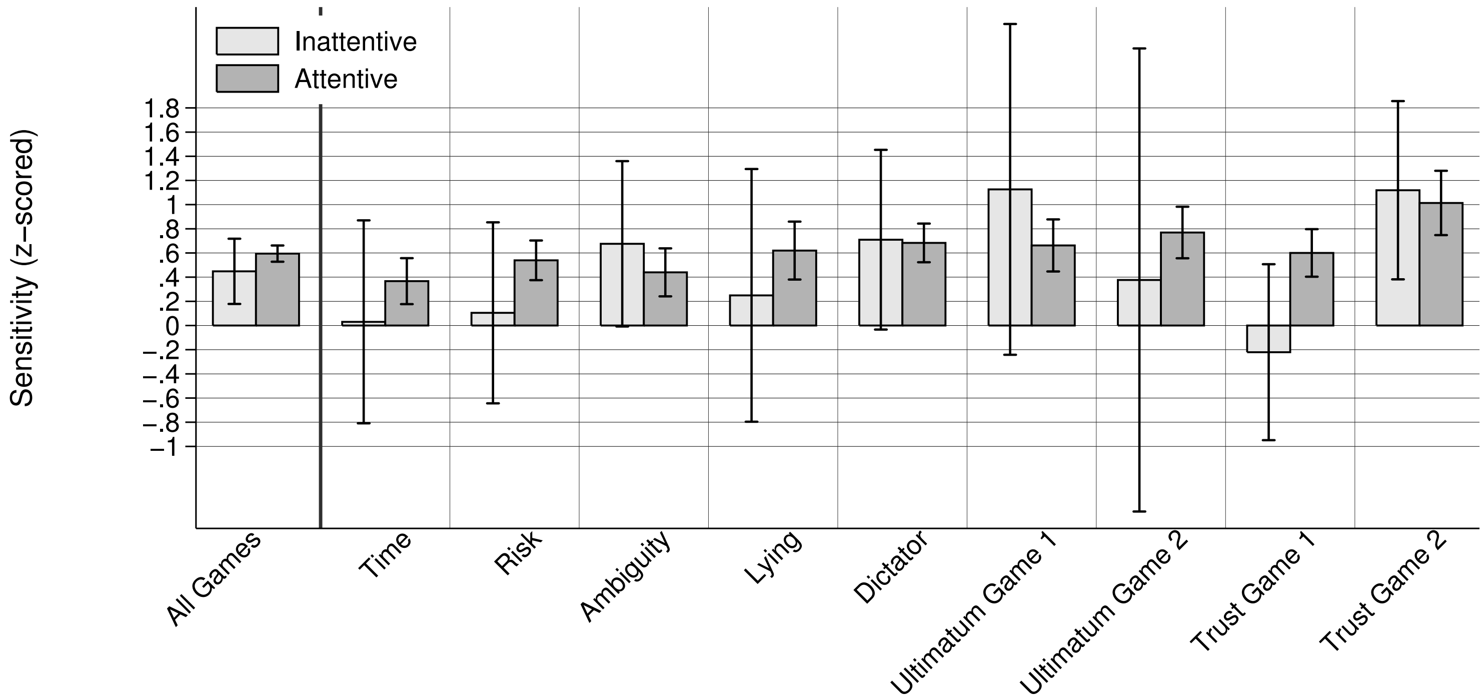
Notes: This figure utilizes MTurk data from experiment 1 and displays the sensitivity of behavior to our strong demand treatments by whether choices are incentivized or hypothetical. In these demand treatments we reveal the experimental objective to our respondents. The behavior in these treatment arms is z-scored at the game-incentive level using the mean and standard deviation in the negative demand condition.

Figure A.2: Gender differences in response to strong demand treatments



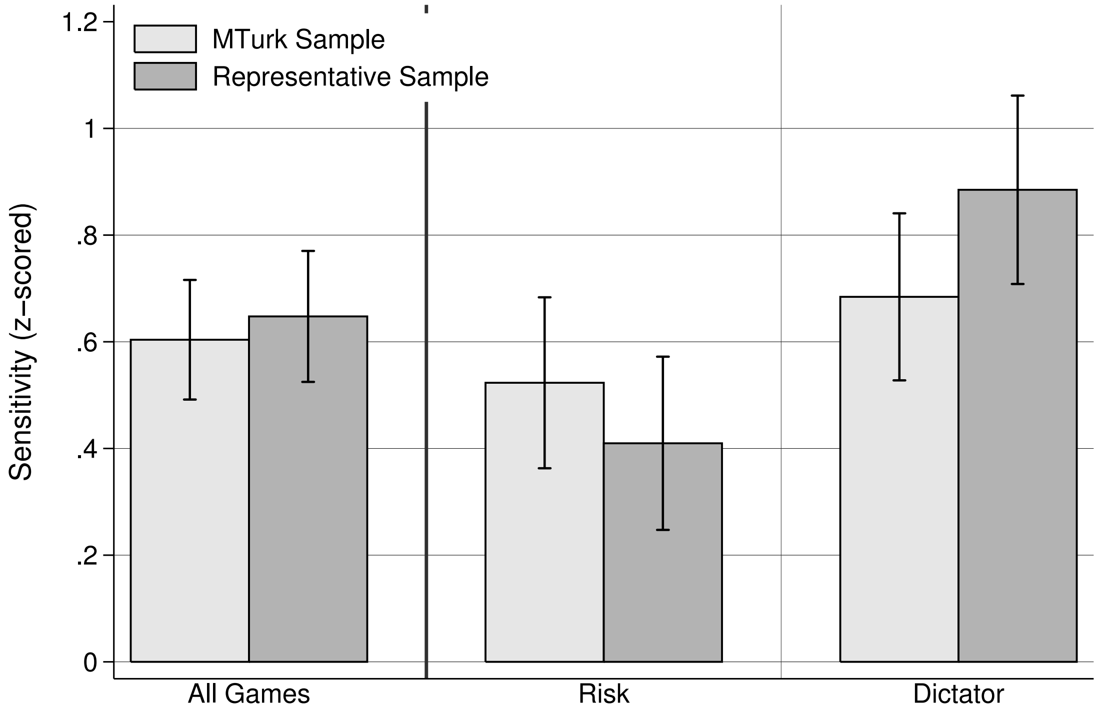
Notes: This figure utilizes data from all MTurk experiments with strong demand treatments using real stakes. This figure displays the sensitivity of behavior to our strong demand treatments for males and females separately across 11 standard experimental paradigms. The behavior in these treatment arms is z-scored at the game-level using the mean and standard deviation in the negative demand condition. In the figure we display the average sensitivity at the game level along with the 95 percent confidence interval. In these demand treatments we reveal the experimental objective to our respondents.

Figure A.3: Response to strong demand treatments by attention



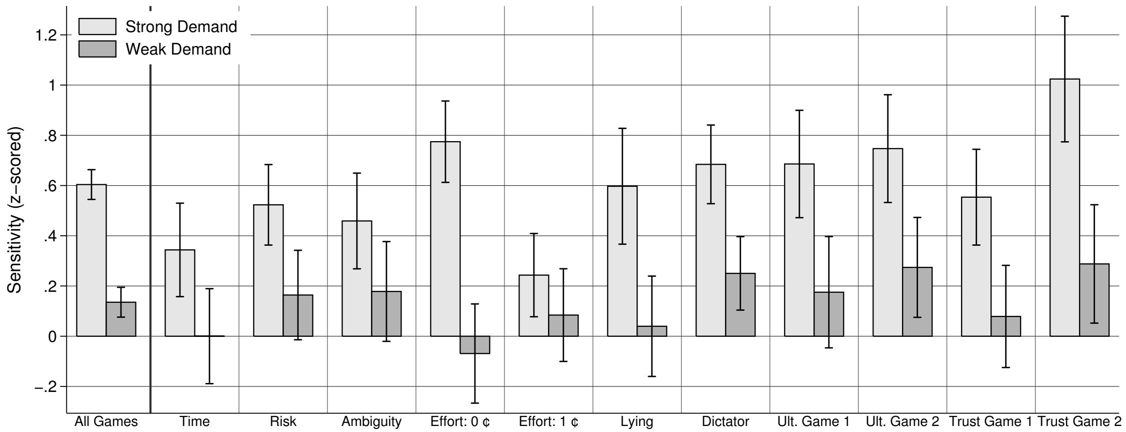
Notes: This figure utilizes data from all MTurk experiments with strong demand treatments using real stakes. This figure displays the response to our strong demand treatments by our respondents' level of attention. The behavior in these treatment arms is z-scored at the game-level using the mean and standard deviation in the negative demand condition. In the figure we display the average sensitivity at the game level along with the 95 percent confidence interval of the sensitivity. In these demand treatments we reveal the experimental objective to our respondents.

Figure A.4: Response to strong demand treatments by population



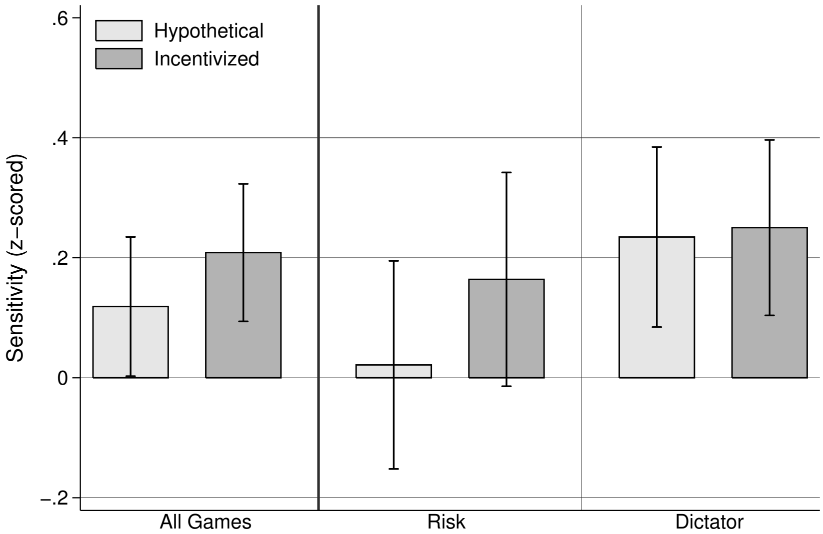
Notes: This figure utilizes data from experiment 1 on MTurk using real stakes as well as data from experiment 4 with the representative online panel. This figure displays the response to our strong demand treatments separately for the MTurk sample and the representative online sample. In the figure we display the average sensitivity at the game level along with the 95 percent confidence interval of the sensitivity. In these demand treatments we reveal the experimental objective to our respondents.

Figure A.5: Response to strong vs. weak demand treatments



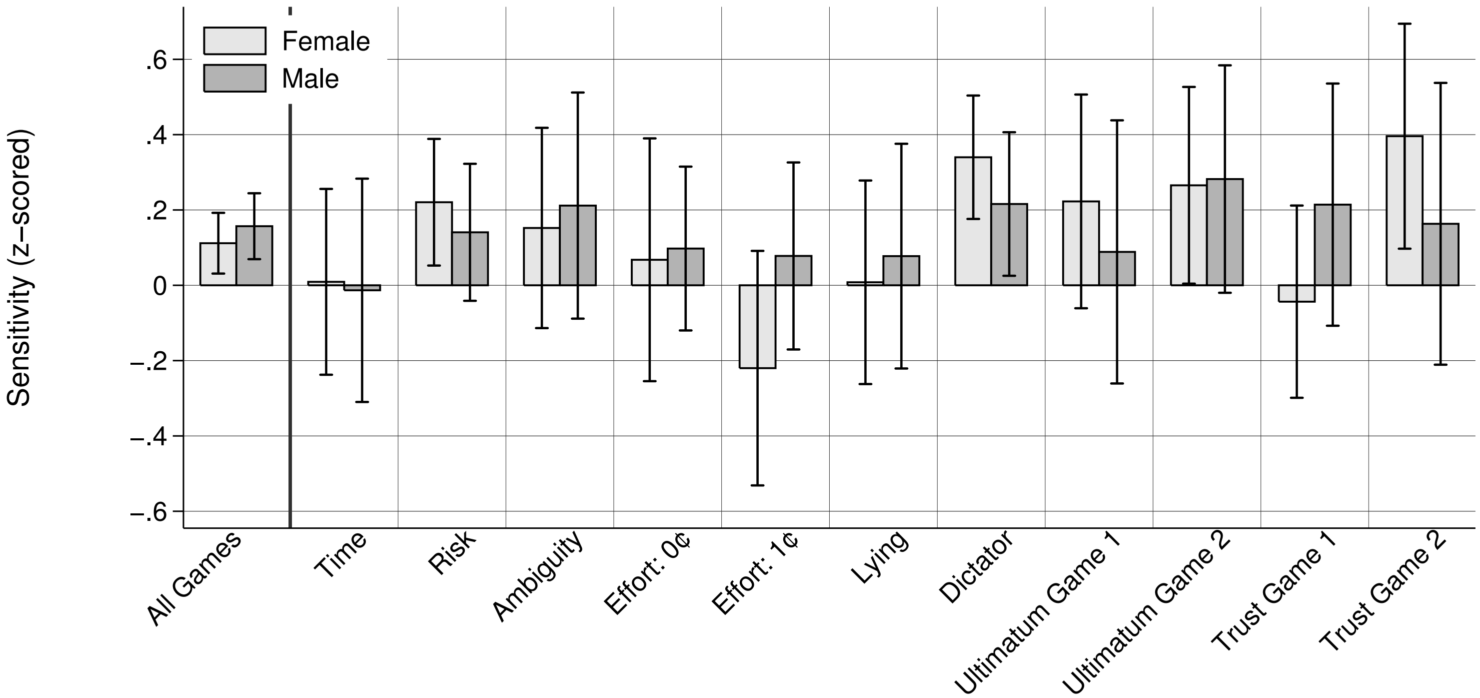
Notes: This figure displays the response to our weak and strong demand treatments separately for the incentivized sample. We display the average sensitivity along with the 95 percent confidence interval of the sensitivity. In these demand treatments we reveal the experimental objective and hypothesis respectively.

Figure A.6: Response to weak demand treatments by Incentives



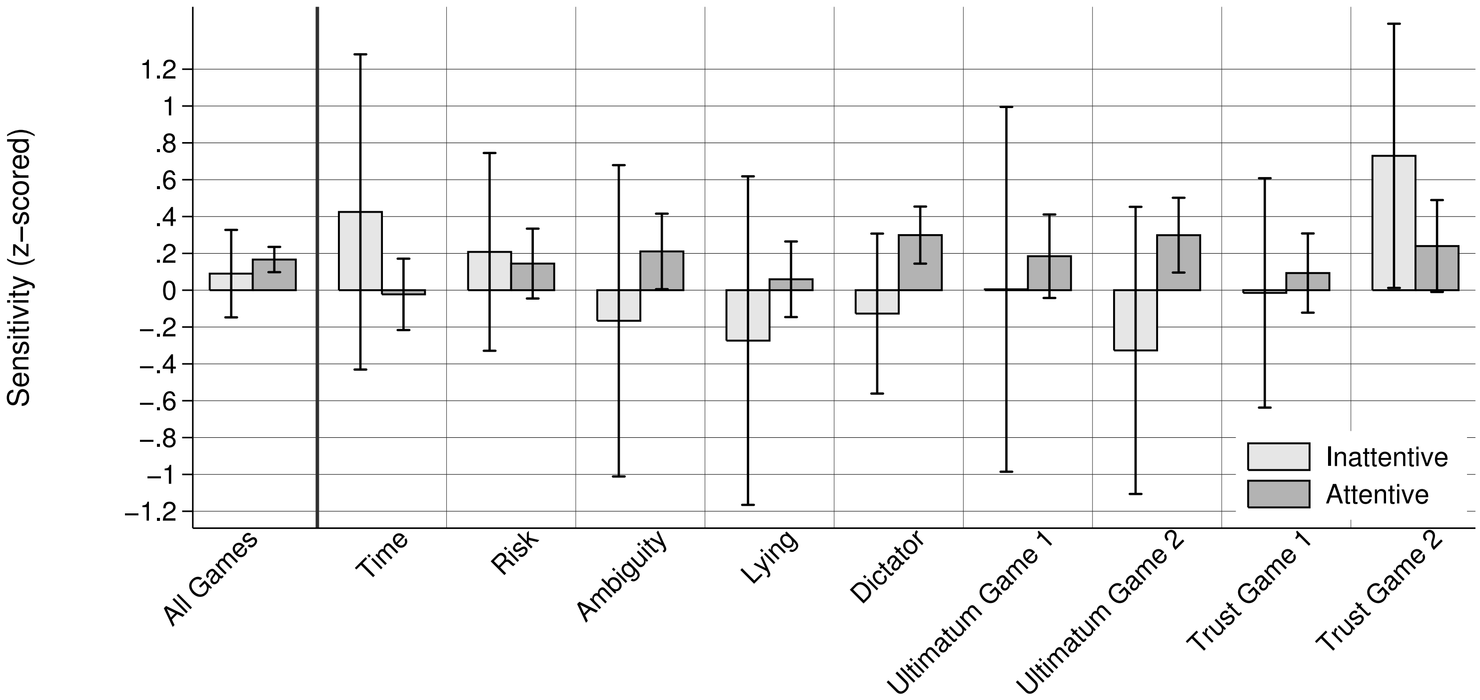
Notes: This figure displays the response to our weak demand treatments separately for the incentivized sample and the sample completing hypothetical choices. We display the average sensitivity along with the 95 percent confidence interval of the sensitivity. In these demand treatments we reveal the experimental hypothesis to our respondents.

Figure A.7: Gender Differences in response to weak demand treatments



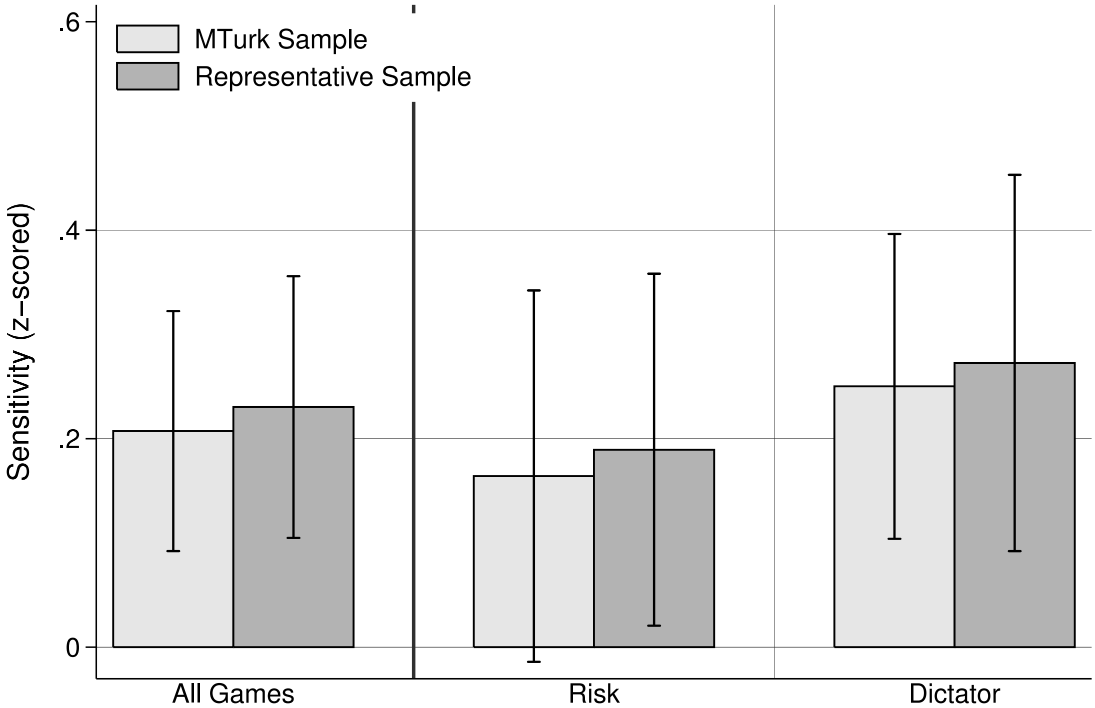
Notes: This figure displays the sensitivity to our weak demand treatments for males and females separately. We display the average sensitivity along with the 95 percent confidence interval of the sensitivity. In these demand treatments we reveal the experimental hypothesis to our respondents.

Figure A.8: Response to weak demand treatments by attention



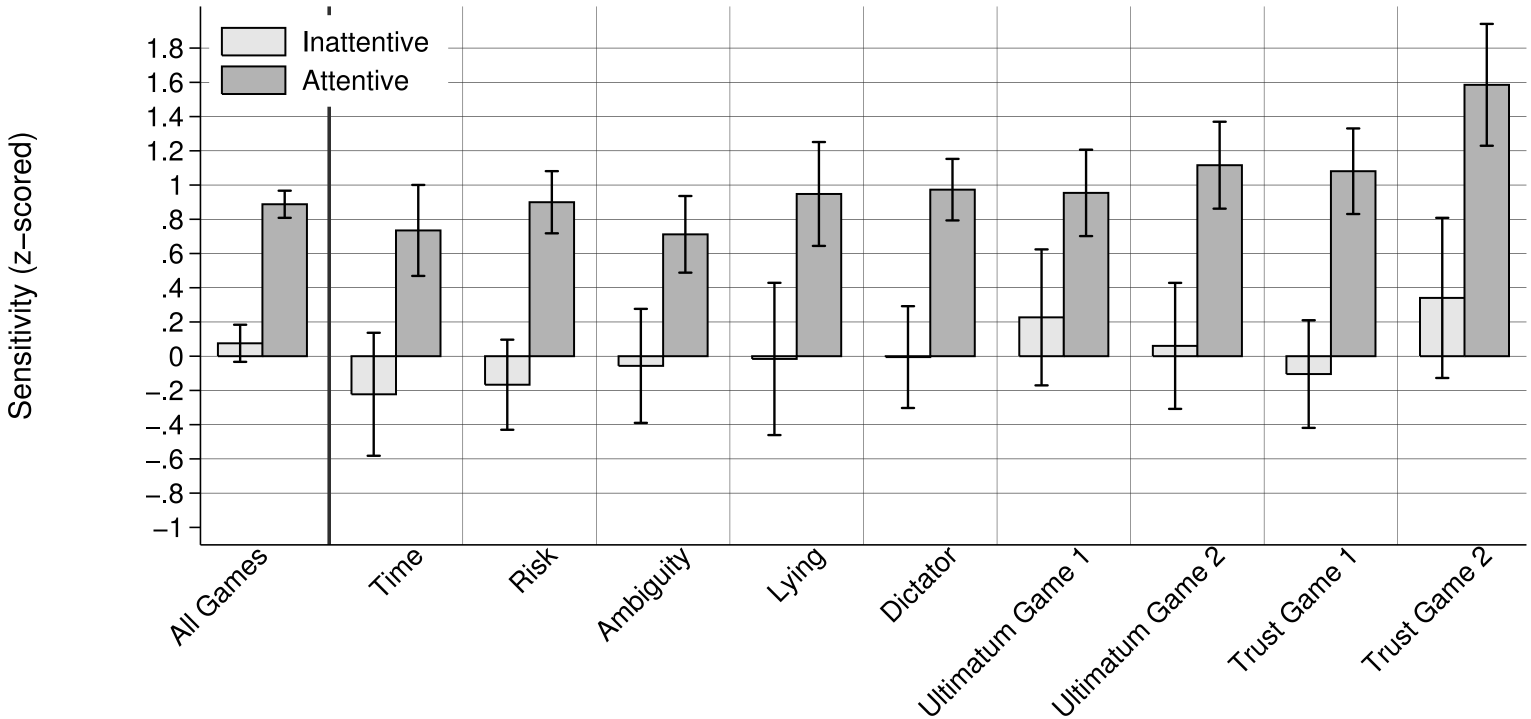
Notes: This figure displays the response to our weak demand treatments by our respondents' level of attention. We display the average sensitivity along with the 95 percent confidence interval of the sensitivity.In these demand treatments we reveal the experimental hypothesis to our respondents.

Figure A.9: Response to weak demand treatments by population



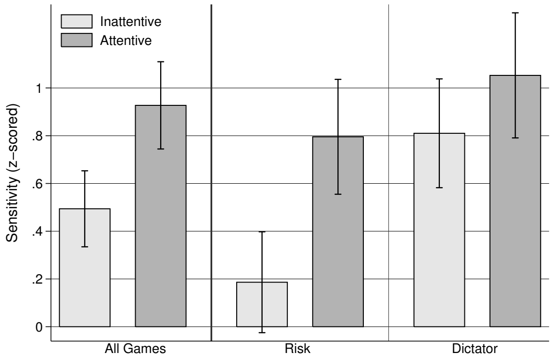
Notes: This figure displays the response to our weak demand treatments separately for the MTurk sample and the representative online sample. We display the average sensitivity at the game level along with the 95 percent confidence interval of the sensitivity. In these demand treatments we reveal the experimental hypothesis to our respondents.

Figure A.10: Response to strong demand treatments by attention (alternative measure)



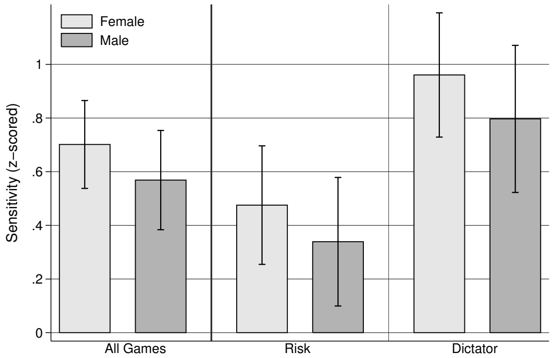
Notes: This figure displays the response to our strong demand treatments separately for attentive and inattentive subjects. Attention takes value one for respondents in the positive demand condition who thought the experimenter wanted a high action or for respondents who are in the negative demand condition and thought that the experimenter wanted a low value.

Figure A.11: Response to weak demand treatments by population



Notes: This figure utilized data from the representative online panel using the strong demand treatments. It displays the response to our weak demand treatments separately for attentive and inattentive subjects. Attentive subjects are those who pass the attention screener.

Figure A.12: Response to weak demand treatments by population



Notes: This figure utilized data from the representative online panel using the strong demand treatments. It displays the response to our weak demand treatments separately for attentive and inattentive subjects. Attentive subjects are those who pass the attention screener.

C Online Appendix: Theoretical Appendix

C.1 Derivation of $E[h|h^L]$

We suppress dependence on ζ to reduce clutter. After observing h^L, the decision-maker's posterior $E[h|h^L]$ equals $Pr(h = 1|h^L) \times 1 + Pr(h = -1|h^L) \times (-1)$ or

$$\begin{aligned}
E[h|h^L = y] &= Pr(h = 1|h^L = y) - Pr(h = -1|h^L = y) \\
&= \frac{A}{B} \\
A &= Pr(h^L = y|h = 1)Pr(h = 1) - Pr(h^L = y|h = -1)Pr(h = -1) \\
B &= Pr(h^L = y|h = 1)Pr(h = 1) + Pr(h^L = y|h = -1)Pr(h = -1)
\end{aligned}$$

Since $Pr(h = j|h^L = y) = \frac{1}{2}(1 - p^L) + p^L\mathbb{I}[y = j]$ and $Pr(h = j) = \frac{1}{2}$ we have

$$\begin{aligned}
A &= \frac{1}{2}\left[\left(\frac{1}{2}(1 - p^L) + p^L\mathbb{I}[y = 1]\right) - \left(\frac{1}{2}(1 - p^L) + p^L\mathbb{I}[y = -1]\right)\right] \\
&= \frac{1}{2}p^L\left(\mathbb{I}[y = 1] - \mathbb{I}[y = -1]\right) = \frac{1}{2}p^L h^L \\
B &= \frac{1}{2}\left[\left(\frac{1}{2}(1 - p^L) + p^L\mathbb{I}[y = 1]\right) + \left(\frac{1}{2}(1 - p^L) + p^L\mathbb{I}[y = -1]\right)\right] \\
&= \frac{1}{2}
\end{aligned}$$

Therefore we can write:

$$E[h|h^L(\zeta)] = h^L(\zeta)p^L(\zeta) \tag{16}$$

$$Pr(h = 1|h^L(\zeta)) = 0.5(1 + h^L p^L(\zeta)) \tag{17}$$

where the latter follows from the fact that $E[h|h^L(\zeta)] = 2Pr(h = 1|h^L(\zeta)) - 1$.

C.2 Derivation of $E[h|h^T, h^L]$

We have assumed that when $h^T = \emptyset$, $E[h|h^T, h^L] = E[h|h^L]$. After observing $h^T \neq \emptyset$, the subject forms a posterior:

$$
\begin{aligned}
E[h|h^T, h^L] &= Pr(h = 1|h^T, h^L) - Pr(h = -1|h^T, h^L) \\
&= \frac{A}{B} \\
A &= Pr(h^T = x|h = 1, h^L = y)Pr(h = 1|h^L = y) \\
&\quad - Pr(h^T = x|h = -1, h^L = y)Pr(h = -1|h^L = y) \\
B &= Pr(h^T = x|h = 1, h^L = y)Pr(h = 1|h^L = y) \\
&\quad + Pr(h^T = x|h = -1, h^L = y)Pr(h = -1|h^L = y)
\end{aligned}
$$

Using the following

$$
\begin{aligned}
Pr(h^T = x|h = j, h^L = y) &= \frac{1}{2}(1 - p^T) + p^T \mathbb{I}[x = j] \\
Pr(h = j|h^L = y) &= \frac{1}{2}(1 - p^L) + p^L \mathbb{I}[y = j]
\end{aligned}
$$

we have:

$$
\begin{aligned}
A = &\left(\frac{1}{2}(1-p^T) + p^T\mathbb{I}[x=1]\right)\left(\frac{1}{2}(1-p^L) + p^L\mathbb{I}[y=1]\right) \\
&- \left(\frac{1}{2}(1-p^T) + p^T\mathbb{I}[x=-1]\right)\left(\frac{1}{2}(1-p^L) + p^L\mathbb{I}[y=-1]\right) \\
&= \frac{1}{2}(1-p^T)p^L(\mathbb{I}[y=1] - \mathbb{I}[y=-1]) \\
&+ \frac{1}{2}(1-p^L)p^T(\mathbb{I}[x=1] - \mathbb{I}[x=-1]) \\
&+ p^Tp^L(\mathbb{I}[x=1]\mathbb{I}[y=1] - \mathbb{I}[x=-1]\mathbb{I}[x=-1]) \\
&= \frac{1}{2}(1-p^T)p^Lh^L + \frac{1}{2}(1-p^L)p^Th^T \\
&+ p^Tp^L(\mathbb{I}[x=1]\mathbb{I}[y=1] - (1-\mathbb{I}[x=1])(1-\mathbb{I}[y=1])) \\
&= \frac{1}{2}(1-p^T)p^Lh^L + \frac{1}{2}(1-p^L)p^Th^T \\
&+ p^Tp^L\left(\mathbb{I}[x=1] - \frac{1}{2}\right) + p^Tp^L\left(\mathbb{I}[y=1] - \frac{1}{2}\right) \\
&= \frac{1}{2}(1-p^T)p^Lh^L + \frac{1}{2}(1-p^L)p^Th^T + \frac{1}{2}p^Tp^L(h^T+h^L) \\
&= \frac{1}{2}\left(p^Th^T + p^Lh^L\right)
\end{aligned}
$$

$$
\begin{aligned}
B &= \left(\frac{1}{2}(1-p^T) + p^T\mathbb{I}[x=1]\right)\left(\frac{1}{2}(1-p^L) + p^L\mathbb{I}[y=1]\right) \\
&+ \left(\frac{1}{2}(1-p^T) + p^T\mathbb{I}[x=-1]\right)\left(\frac{1}{2}(1-p^L) + p^L\mathbb{I}[y=-1]\right) \\
&= \frac{1}{2}(1-p^T)(1-p^L) + \frac{1}{2}(1-p^T)p^L(\mathbb{I}[y=1] + \mathbb{I}[y=-1]) \\
&+ \frac{1}{2}(1-p^L)p^T(\mathbb{I}[x=1] + \mathbb{I}[x=-1]) \\
&+ p^Tp^L(\mathbb{I}[x=1]\mathbb{I}[y=1] + \mathbb{I}[x=-1]\mathbb{I}[y=-1]) \\
&= \frac{1}{2}(1-p^Tp^L) + 2p^Tp^L\mathbb{I}[x=1]\mathbb{I}[y=1] \\
&+ p^Tp^L\left[\left(\frac{1}{2} - \mathbb{I}[x=1]\right) + \left(\frac{1}{2} - \mathbb{I}[y=1]\right)\right] \\
&= \frac{1}{2}(1-p^Tp^L) + 2p^Tp^L\left(\frac{1}{2}(h^T+1)\right)\left(\frac{1}{2}(h^L+1)\right) - \frac{1}{2}p^Tp^L(h^T+h^L) \\
&= \frac{1}{2}(1+p^Th^Tp^Lh^L)
\end{aligned}
$$

Which uses the facts that $\mathbb{I}[x=1] - \mathbb{I}[x=-1] = h^T$, $\mathbb{I}[x=1] + \mathbb{I}[x=-1] = 1$, and $\mathbb{I}[x=1] = \frac{1}{2}(h^T+1)$. Therefore we can write:

$$
E[h|h^T, h^L(\zeta)] = \frac{h^L(\zeta)p^L(\zeta) + h^Tp^T}{1 + h^L(\zeta)p^L(\zeta)h^Tp^T} \tag{18}
$$

$$
Pr(h=1|h^T, h^L(\zeta)) = 0.5\left(1 + \frac{h^L(\zeta)p^L(\zeta) + h^Tp^T}{1 + h^L(\zeta)p^L(\zeta)h^Tp^T}\right) \tag{19}
$$

where the latter follows from the fact that $E[h|h^T, h^L(\zeta)] = 2Pr(h = 1|h^T, h^L(\zeta)) - 1$.

C.3 Proof of Proposition 1 (Monotone demand treatment effects)

We are interested in the sign of $\phi(E[h|h^T, h^L(\zeta)] - E[h|h^L(\zeta)])$. We have:

$$
\begin{aligned}
\phi(E[h|h^T, h^L(\zeta)] - E[h|h^L(\zeta)]) &= \phi\left(\frac{h^L(\zeta)p^L(\zeta) + h^T p^T}{1 + h^L(\zeta)p^L(\zeta)h^T p^T} - h^L(\zeta)p^L(\zeta)\right) \\
&= \phi h^T p^T \frac{(1 - h^L(\zeta)^2 p^L(\zeta)^2)}{1 + h^L(\zeta)p^L(\zeta)h^T p^T}
\end{aligned}
$$

Because we assumed that $p^L(\zeta) < 1$, this expression has the same sign as $\phi h^T p^T$. We want to show that $\phi(E[h|h^T = 1, h^L(\zeta)] - E[h|h^L(\zeta)]) \geq 0$ and $\phi(E[h|h^T = -1, h^L(\zeta)] - E[h|h^L(\zeta)]) \leq 0$. This follows trivially when $p^T = 0$. When $p^T > 0$ if follows if and only if $\phi \geq 0$.

C.4 Conditions for Monotone Sensitivity

Assumption 3 (monotone sensitivity) assumes that sensitivity $S(\zeta) = a^+(\zeta) - a^-(\zeta)$ is (strictly) monotone in the size of the latent demand effect $\left|a^L(\zeta) - a(\zeta)\right|$. Here we examine cases under which that is and is not the case. We assume throughout that Assumptions 1 and 2 hold.

C.4.1 Variation driven by ϕ.

We are interested in how ϕ affects latent demand ($d\left|a^L(\zeta) - a(\zeta)\right|/d\phi$) and sensitivity ($dS(\zeta)/d\phi$). From (5) we obtain:

$$
\frac{d(a^L(\zeta) - a(\zeta))}{d\phi} = -\frac{h^L(\zeta)p^L(\zeta)}{v_{11}(a^L(\zeta), \zeta)}
$$

which has the same sign as $h^L(\zeta)$, allowing us to write $\frac{d\left|a^L(\zeta) - a(\zeta)\right|}{d\phi} = -\frac{p^L(\zeta)}{v_{11}(a^L(\zeta),\zeta)} \geq 0$.

Turning to sensitivity, we have:

$$
\begin{aligned}
\frac{dS(\zeta)}{d\phi} &= \frac{da^+(\zeta)}{d\phi} - \frac{da^-(\zeta)}{d\phi} \\
&= -\frac{1}{v_{11}(a^+(\zeta), \zeta)} \frac{h^L(\zeta)p^L(\zeta) + p^T}{1 + h^L(\zeta)p^L(\zeta)p^T} + \frac{1}{v_{11}(a^-(\zeta), \zeta)} \frac{h^L(\zeta)p^L(\zeta) - p^T}{1 - h^L(\zeta)p^L(\zeta)p^T}
\end{aligned}
$$

By Assumption 2, $h^L(\zeta)p^L(\zeta)+p^T \geq 0$ and $h^L(\zeta)p^L(\zeta)+p^T \leq 0$, so both terms are positive, i.e. $\frac{dS(\zeta)}{d\phi} \geq 0$. Therefore Monotone Sensitivity holds and any set of environments that differ only in ϕ constitutes a comparison class, i.e. for such environments, sensitivity is informative about the magnitude of latent demand effects.

Example 3. Suppose subject pool A is more concerned for pleasing the experimenter than subject pool B. Then latent demand effects and sensitivity will be larger in magnitude in subject pool A.

C.4.2 Variation driven by v.

Suppose that ζ can be separated into a parameter, z, and a remainder term, ζ', that v is differentiable in z and that ϕ, h^L and p^L do not depend on z. z could be a preference parameter (e.g. risk aversion) or a design parameter (e.g. the scale of incentives). We write $U(a,\zeta',z) = v(a,\zeta',z) + a\phi(\zeta')E[h|\zeta']$ and modify the first-order conditions accordingly.

$$
\begin{aligned}
\frac{d(a^L(\zeta',z)-a(\zeta',z))}{dz} &= \frac{da^L(\zeta',z)}{dz} - \frac{da(\zeta',z)}{dz} \\
&= -\left[\frac{v_{13}(a^L(\zeta',z),\zeta',z)}{v_{11}(a^L(\zeta',z),\zeta',z)} - \frac{v_{13}(a(\zeta',z),\zeta',z)}{v_{11}(a(\zeta',z),\zeta',z)}\right] \\
\frac{dS(\zeta',z)}{dz} &= -\left[\frac{v_{13}(a^+(\zeta',z),\zeta',z)}{v_{11}(a^+(\zeta',z),\zeta',z)} - \frac{v_{13}(a^-(\zeta',z),\zeta',z)}{v_{11}(a^-(\zeta',z),\zeta',z)}\right]
\end{aligned}
$$

It is clear from inspecting these conditions that we need to know how v_{13}/v_{11} varies with a, i.e.:

$$
\frac{d\frac{v_{13}(a,\zeta',z)}{v_{11}(a,\zeta',z)}}{da} = \frac{v_{11}(a,\zeta',z)v_{113}(a,\zeta',z) - v_{111}(a,\zeta',z)v_{13}(a,\zeta',z)}{v_{11}(a,\zeta',tz)}
$$

It is difficult to make general statements about these objects for general utility functions, so we focus attention on two special cases of interest.

Multiplicative separability. Suppose that $v(a, \zeta', z) = \nu(a, \zeta')f(z)$ and define z such that $f'(z) > 0$. Then

$$\begin{aligned}\frac{d\left(a^L(\zeta', z) - a(\zeta', z)\right)}{dz} &= -f'(z)\left[\frac{\nu_1(a^L(\zeta', z), \zeta')}{\nu_{11}(a^L(\zeta', z), \zeta')} - \frac{\nu_1(a(\zeta', z), \zeta')}{\nu_{11}(a(\zeta', z), \zeta')}\right] \\ &= -f'(z)\frac{\nu_1(a^L(\zeta', z), \zeta')}{\nu_{11}(a^L(\zeta', z), \zeta')}\end{aligned}$$

Since by concavity $\nu_1(a, \zeta') > 0$ for $a < a(\zeta', z)$ and $\nu_1(a, \zeta') < 0$ for $a > a(\zeta', z)$, we have $\frac{d\left|a^L(\zeta', z) - a(\zeta', z)\right|}{dz} \leq 0$. Similarly

$$\frac{dS(\zeta)}{dz} = -f'(z)\left[\frac{\nu_1(a^+(\zeta', z), \zeta')}{\nu_{11}(a^+(\zeta', z), \zeta')} - \frac{\nu_1(a^-(\zeta', z), \zeta')}{\nu_{11}(a^-(\zeta', z), \zeta')}\right]$$

Since $\nu_1(a^+(\zeta', z), \zeta') \leq 0$ and $\nu_1(a^-(\zeta', z), \zeta') \geq 0$, we have $\frac{dS(\zeta)}{dz} \leq 0$. Therefore Monotone Sensitivity holds and any set of environments that varies only in z is a valid comparison set.

Intuitively, this case captures changes in the slope of payoffs that leave the optimal natural action unchanged. For example, an increase in the scale of incentives that makes the payoff function "more concave" around the natural action makes deviating from the natural action more costly and so decreases the magnitude of latent demand and sensitivity.

Example 4 (Belief scoring)**.** Consider a belief-reporting task rewarded by a quadratic scoring rule. A risk-neutral subject reports a belief, a, which is the probability of an event A. He is paid $\frac{z}{2}\,[1-(\mathbb{I}[A]-a)^2]$ where $\mathbb{I}[A] = 1$ if A is true and 0 otherwise. The utility function is $U(a, \zeta', z) = \frac{z}{2}\,[1-\mu(1-a)^2 - (1-\mu)(-a)^2] + a\phi(\zeta')E[h|\zeta']$, so $f(z) = z$. The optimal action solves $z\,[\mu(1-a^*) - (1-\mu)a^*] + \phi(\zeta')E[h|\zeta'] = 0$ or $a^* = \mu + \frac{\phi(\zeta')E[h|\zeta']}{z}$. Increases in z are equivalent to decreases in ϕ and decrease both the magnitude of latent demand effects, and sensitivity.

Additive separability. Suppose that $v(a, \zeta', z) = v(a, \zeta') + af(z)$ and define z such that $f'(z) > 0$. Then:

$$\frac{d(a^L(\zeta', z) - a(\zeta', z))}{dz} = -f'(z) \left[\frac{1}{\nu_{11}(a^L(\zeta', z), \zeta')} - \frac{1}{\nu_{11}(a(\zeta', z), \zeta')} \right]$$

and

$$\frac{dS(\zeta)}{dz} = -f'(z) \left[\frac{1}{\nu_{11}(a^+(\zeta', z), \zeta')} - \frac{1}{\nu_{11}(a^-(\zeta', z), \zeta')} \right]$$

What matters in this case is how the concavity of v (and therefore ν) with respect to a varies with a. Suppose $\nu_{111} < 0$, so ν_{11} is decreasing in a, i.e. concavity is increasing. Then $\frac{dS(\zeta)}{dz} < 0$, i.e. increases in z decrease sensitivity. If $a^L(\zeta', z) > a(\zeta', z)$ then $\frac{d(a^L(\zeta',z) - a(\zeta',z))}{dz} < 0$ and if $a^L(\zeta', z) < a(\zeta', z)$ then $\frac{d(a^L(\zeta',z) - a(\zeta',z))}{dz} > 0$, so $\frac{d\left|a^L(\zeta',z) - a(\zeta',z)\right|}{dz} < 0$ and Monotone Sensitivity holds. Monotone sensitivity also holds (with the inequalities reversed) for $\nu_{111} > 0$.

Example 5 (Effort provision)**.** A subject performs a real-effort task for piece rate z with cost of effort $C(a)$, $C' > 0$,$C'' > 0$, $C''' > 0$. $U(a, \zeta', z) = za - C(a) + a\phi(\zeta')E[h|\zeta']$. The optimal action a^* solves $z - C'(a^*) + \phi(\zeta')E[h|\zeta'] = 0$. As z increases, a^* increases and responsiveness to latent demand or demand treatments decreases.

C.4.3 Variation driven by inattention.

Suppose that with some probability ξ the subject is an attentive type who pays careful attention to the decision-making environment, and with probability $1 - \xi$, he is inattentive. When inattentive, he takes some action $a^I(\zeta)$. $a^I(\zeta)$ might be equal to $a(\zeta)$, in which case the subject is only inattentive to experimenter demand, but it might differ if the subject is also inattentive to other design features.

While until now we have treated the actions as those of a representative agent, for this analysis it is more natural to work with expected or average actions over a sample. Denote by $\bar{a}(\zeta) = \xi a(\zeta) + (1-\xi)a^I(\zeta)$ the expected natural action, define $\bar{a}^L(\zeta), \bar{a}^+(\zeta), \bar{a}^-(\zeta)$ equivalently and let $\bar{S}(\zeta) = \bar{a}^+(\zeta) - \bar{a}^-(\zeta)$. The latent demand effect is now equal to $\left|\bar{a}^L(\zeta) - \bar{a}(\zeta)\right| = \xi \left|a^L(\zeta) - a(\zeta)\right|$, while $\bar{S}(\zeta) = \xi S(\zeta)$. Hence, if the variation in latent demand effects is driven by variation in attention, ξ, Monotone Sensitivity will hold, and any set of environments that varies only in subject attentiveness is a valid comparison set. Note that since we have assumed the subject is inattentive to both latent demand and the demand treatment, bounding will hold if $p^T \geq p^L$ as before.

C.4.4 Variation driven by beliefs.

Consider changes to the environment that influence behavior only by altering participants' beliefs about the experimenter's objective, i.e. we consider variation in $h^L(\zeta)p^L(\zeta)$. Call this term H. $a(\zeta)$ is unaffected, so:

$$\frac{d(a^L(\zeta) - a(\zeta))}{dH} = -\frac{\phi(\zeta)}{v_{11}(a^L(\zeta), \zeta)} \geq 0$$

and therefore $\frac{d\left|a^L(\zeta)-a(\zeta)\right|}{dH} = -\frac{\phi(\zeta)}{v_{11}(a^L(\zeta),\zeta)} \times \text{sign}(a^L(\zeta) - a(\zeta)) = -\frac{\phi(\zeta)h^L(\zeta)}{v_{11}(a^L(\zeta),\zeta)}$ which is positive when $h^L(\zeta) = 1$ (because an increase in H means the subject's beliefs are shifting toward certainty that the experimenter wants a high action) and negative when $h^L(\zeta) = -1$ (because the subject is becoming more uncertain about the experimenter's wishes).

Next we turn to demand treatment effects. First we derive the response of the subject's posterior:

$$\frac{d\frac{H+h^T p^T}{1+Hh^T p^T}}{dH} = \frac{\left(1+Hh^T p^T\right) - \left(H+h^T p^T\right) h^T p^T}{\left(1+Hh^T p^T\right)^2}$$
$$= \frac{1-\left(h^T p^T\right)^2}{\left(1+Hh^T p^T\right)^2} = \frac{1-p^{T2}}{\left(1+Hh^T p^T\right)^2}$$

So:

$$\frac{dS(\zeta)}{dH} = -\phi(\zeta)(1-p^{T2})\left[\frac{1}{\left(1+Hp^T\right)^2 v_{11}(a^+(\zeta),\zeta)} - \frac{1}{\left(1-Hp^T\right)^2 v_{11}(a^-(\zeta),\zeta)}\right]$$

The sign of this expression depends on the sign of H and how v_{11} changes with a. However, it is straightforward to see that Monotone Sensitivity *will not* hold in general, and in fact sensitivity will tend to be higher when latent demand is weaker. To see this, consider the simple case where v_{11} is constant. Then we have:

$$\frac{dS(\zeta)}{dH} = -\frac{\phi(\zeta)(1-p^{T2})}{v_{11}}\left[\frac{\left(1-Hp^T\right)^2 - \left(1+Hp^T\right)^2}{\left(1+Hp^T\right)^2\left(1-Hp^T\right)^2}\right]$$
$$= -\frac{\phi(\zeta)(1-p^{T2})}{v_{11}}\left[\frac{-4Hp^T}{\left(1+Hp^T\right)^2\left(1-Hp^T\right)^2}\right]$$

which is positive when $h^L = -1$ and negative when $h^L = 1$, i.e. it has the opposite sign to $\frac{d\left|a^L(\zeta)-a(\zeta)\right|}{dH}$. The reason is that as H approaches zero, the subject becomes more uncertain about the experimenter's wishes and is therefore very responsive to the new information in the demand treatments. Meanwhile as H approaches 1 or -1, the subject is very confident about the value of h. Although his confidence can be undermined by a demand treatment in the opposite direction, he responds little to a demand treatment that confirms his beliefs, so sensitivity is low.

C.5 Extension: learning about ϕ

An interpretation of our demand treatments is that they signal not only the direction of the experimenter's objective, but the salience or intensity of her preference over objectives. For instance "do me a favor" suggests that the choice is important. We now assume that the decision-maker's preferences are:

$$U(a, \zeta) = v(a, \zeta) + a\phi(\zeta)E[gh|\zeta]$$

where $g \in \{0, 1\}$ captures whether conforming to h is important (1) or unimportant (0) to the experimenter. ϕ remains the decision-maker's preference for pleasing the experimenter, which is now scaled by g, i.e. the decision-maker internalizes the perceived importance of the objective. We assume that g and h are believed independent (i.e. direction and importance are independent), so $E[gh|\zeta] = E[g|\zeta]E[h|\zeta]$. We also assume for simplicity is that the decision-maker's prior $E[g] = 0.5$.

Now, ζ contains two signals, $h^L(\zeta)$, defined as before, and $g^L(\zeta) \in \{0, 1\}$, where $E[g|g^L(\zeta)] = E[g|g^L(\zeta), \zeta]$ (i.e. g^L is a sufficient statistic). g^L is believed to equal g with probability $q^L(\zeta) < 1$ and pure independent noise otherwise. We show below that $E[g|g^L(\zeta)] = \frac{1}{2} + q^L\left(g^L - \frac{1}{2}\right)$.

Similarly, a demand treatment is now two signals (h^T, g^T), where h^T is defined as before and $g^T \in \{0, 1, \emptyset\}$. $g^T = \emptyset$ corresponds to the case where no treatment is used, $g^T = 0$ signals to the subject that their action is not important to the experimenter, and $g^T = 1$ signals that it is.

Conditional on sending a demand treatment, g^T is believed to equal g with probability q^T and otherwise be pure noise independent of all other signals. We show below that the Bayesian posterior is:

$$E[g|g^T, g^L(\zeta)] = \frac{\frac{1}{2} + q^L(\zeta)\left(g^L(\zeta) - \frac{1}{2}\right) + q^T\left(g^T - \frac{1}{2}\right) + q^T q^L(\zeta)\left(\mathbb{I}[g^T = g^L(\zeta)] - \frac{1}{2}\right)}{1 + 2q^T q^L(\zeta)\left(\mathbb{I}[g^T = g^L(\zeta)] - \frac{1}{2}\right)}$$

We assume that g^T can be varied independently of h^T and will be held constant within a typical pair of positive and negative demand treatments.

For bounding to hold, we now need:

$$\phi(\zeta)E[g|g^T, g^L(\zeta)]E[h|h^T = 0, h^L(\zeta)] \leq 0 \leq \phi(\zeta)E[g|g^T, g^L(\zeta)]E[h|h^T = 1, h^L(\zeta)]$$

Since $E[g|g^T, g^L(\zeta)] \geq 0$ our bounding condition does *not* depend on how the demand treatments affect beliefs about g, all we require is $\phi(\zeta) \geq 0$ and $p^T \geq p^L(\zeta)$ as before.[43]

However, beliefs about g do affect the width of the bounds: sensitivity is increasing in $E[g|g^T, g^L(\zeta)]$. The tightest bounds are obtained when $E[g|g^T, g^L(\zeta)] = 0$, which obtains when $g^T = 0$ and $q^T = 1$. More generally, the bounds are tightened by signaling that acting according to the experimenter's objective is not important ($g^T = 0$), or if $g^T = 1$ by minimizing q^T. We suspect that it may be difficult in practice to both strongly signal the direction of the objective (large p^T), which is required for bounding, and that the objective is not important ($g^T = 0$), so reasonable demand treatments are likely to be those that strongly signal a directional objective while keeping salience low, i.e. large p^T and small q^T with $g^T = 1$.

[43]For monotone demand treatment effects to hold, we require

$$\phi(\zeta)E[g|g^T, g^L(\zeta)]E[h|h^T = 0, h^L(\zeta)] \leq \phi(\zeta)E[g|g^L(\zeta)]E[h|h^L(\zeta)] \leq \phi(\zeta)E[g|g^T, g^L(\zeta)]E[h|h^T = 1, h^L(\zeta)]$$

We can write

$$\phi(\zeta)\frac{E[h|h^T = 0, h^L(\zeta)]}{E[h|h^L(\zeta)]} \leq \phi(\zeta)\frac{E[g|g^L(\zeta)]}{E[g|g^T, g^L(\zeta)]} \leq \phi(\zeta)\frac{E[h|h^T = 1, h^L(\zeta)]}{E[h|h^L(\zeta)]}$$

We see that $\phi(\zeta) \geq 0$ is necessary but not sufficient for monotone demand treatment effects, we also need that $E[g|g^T, g^L(\zeta)]$ is neither "too big" nor "too small" relative to $E[g|g^L(\zeta)]$. Intuitively, if $g^T = 1$ the demand treatments shift all actions further away from the natural action, while if $g^T = 0$. all actions are shifted toward the natural action. $g^T = 1$ and $p^T \geq p^L$ are sufficient for monotone demand treatments to hold.

C.5.1 Derivation of $E[g|g^L(\zeta)]$ and $E[g|g^T, g^L(\zeta)]$

Let the prior belief be $\frac{1}{2}$.

$$\begin{aligned}
E[g|g^L = y] &= Pr(g = 1|g^L = y) \\
&= \frac{A}{B} \\
A &= Pr(g^L = y|g = 1)Pr(g = 1) \\
B &= Pr(g^L = y|g = 1)Pr(g = 1) + Pr(g^L = y|g = 0)Pr(g = 0)
\end{aligned}$$

Since $Pr(g = j|g^L = y) = \frac{1}{2}(1 - q^L) + q^L\mathbb{I}[y = j]$ and $Pr(g = j) = \frac{1}{2}$ we have

$$\begin{aligned}
A &= \frac{1}{2}\left(\frac{1}{2}(1 - q^L) + q^L\mathbb{I}[y = 1]\right) \\
&= \frac{1}{2}\left(\frac{1}{2} + q^L\left(g^L - \frac{1}{2}\right)\right) \\
B &= \frac{1}{2}\left[\left(\frac{1}{2}(1 - q^L) + q^L\mathbb{I}[y = 1]\right) + \left(\frac{1}{2}(1 - q^L) + q^L\mathbb{I}[y = 0]\right)\right] \\
&= \frac{1}{2}
\end{aligned}$$

Therefore, $E[g|g^L(\zeta)] = \frac{1}{2} + q^L\left(g^L - \frac{1}{2}\right)$.

Turning to $E[g|g^T, g^L(\zeta)]$, we have assumed that when $g^T = \emptyset$, $E[g|g^T, g^L] = E[g|g^L]$. After observing $g^T \neq \emptyset$, the subject forms a posterior:

$$
\begin{aligned}
E[g|g^T, g^L] &= Pr(g = 1|g^T, g^L) \\
&= \frac{A}{B} \\
A &= Pr(g^T = x|g = 1, g^L = y)Pr(g = 1|g^L = y) \\
B &= Pr(g^T = x|g = 1, g^L = y)Pr(g = 1|g^L = y) \\
&+ Pr(g^T = x|g = 0, g^L = y)Pr(g = 0|g^L = y)
\end{aligned}
$$

Using the following

$$
\begin{aligned}
Pr(g^T = x|g = j, g^L = y) &= \frac{1}{2}(1 - q^T) + q^T \mathbb{I}[x = j] \\
Pr(g = j|g^L = y) &= \frac{1}{2}(1 - q^L) + q^L \mathbb{I}[y = j]
\end{aligned}
$$

we have:

$$
\begin{aligned}
A &= \left(\frac{1}{2}(1-q^T) + q^T \mathbb{I}[x=1]\right)\left(\frac{1}{2}(1-q^L) + q^L \mathbb{I}[y=1]\right) \\
&= \left(\frac{1}{2}(1-q^T) + q^T g^T\right)\left(\frac{1}{2}(1-q^L) + q^L g^L\right) \\
&= \frac{1}{2}(1-q^T)q^L g^L + \frac{1}{2}(1-q^L)q^T g^T \\
&+ \frac{1}{4}(1-q^T)(1-q^L) + q^T g^T q^L g^L \\
&= \frac{1}{2}(1+q^T(g^T-1))q^L g^L + \frac{1}{2}(1+q^L(g^L-1))q^T g^T \\
&+ \frac{1}{4}(1-q^T)(1-q^L) \\
&= \frac{1}{2}q^L g^L + \frac{1}{2}q^T g^T - \frac{1}{2}q^T q^L \left(g^L(1-g^T) + g^T(1-g^L)\right) \\
&+ \frac{1}{4}(1-q^T)(1-q^L) \\
&= \frac{1}{2}q^L g^L + \frac{1}{2}q^T g^T - \frac{1}{2}q^T q^L \left(\mathbb{I}[g^L \neq g^T]\right) \\
&+ \frac{1}{4} - \frac{1}{4}q^T - \frac{1}{4}q^L + \frac{1}{4}q^T q^L \\
&= \frac{1}{2}q^L\left(g^L - \frac{1}{2}\right) + \frac{1}{2}q^T\left(g^T - \frac{1}{2}\right) - \frac{1}{2}q^T q^L \left(1 - \mathbb{I}[g^L = g^T]\right) \\
&+ \frac{1}{4} + \frac{1}{4}q^T q^L \\
&= \frac{1}{4} + \frac{1}{2}q^L\left(g^L - \frac{1}{2}\right) + \frac{1}{2}q^T\left(g^T - \frac{1}{2}\right) \\
&+ \frac{1}{2}q^T q^L \left(\mathbb{I}[g^T = g^L] - \frac{1}{2}\right)
\end{aligned}
$$

$$
\begin{aligned}
B &= \left(\frac{1}{2}(1-q^T) + q^T\mathbb{I}[x=1]\right)\left(\frac{1}{2}(1-q^L) + q^L\mathbb{I}[y=1]\right) \\
&+ \left(\frac{1}{2}(1-q^T) + q^T\mathbb{I}[x=0]\right)\left(\frac{1}{2}(1-q^L) + q^L\mathbb{I}[y=0]\right) \\
&= \left(\frac{1}{2}(1-q^T) + q^T g^T\right)\left(\frac{1}{2}(1-q^L) + q^L g^L\right) \\
&+ \left(\frac{1}{2}(1-q^T) + q^T(1-g^T)\right)\left(\frac{1}{2}(1-q^L) + q^L(1-g^L)\right) \\
&= \frac{1}{2}(1-q^T)q^L g^L + \frac{1}{2}(1-q^L)q^T g^T \\
&+ \frac{1}{2}(1-q^T)q^L(1-g^L) + \frac{1}{2}(1-q^L)q^T(1-g^T) \\
&+ \frac{1}{2}(1-q^T)(1-q^L) \\
&+ q^T q^L g^T g^L + q^T q^L(1-g^T)(1-g^L) \\
&= \frac{1}{2}(1-q^T)q^L + \frac{1}{2}(1-q^L)q^T + \frac{1}{2}(1-q^T)(1-q^L) \\
&+ q^T q^L \mathbb{I}[g^T = g^L] \\
&= \frac{1}{2} + q^T q^L \left(\mathbb{I}[g^T = g^L] - \frac{1}{2}\right)
\end{aligned}
$$

Therefore,

$$
E[g|g^T, g^L] = \frac{\frac{1}{2} + q^L\left(g^L - \frac{1}{2}\right) + q^T\left(g^T - \frac{1}{2}\right) + q^T q^L\left(\mathbb{I}[g^T = g^L] - \frac{1}{2}\right)}{1 + 2q^T q^L\left(\mathbb{I}[g^T = g^L] - \frac{1}{2}\right)}
$$

C.6 Richer beliefs and correlated signals

In this section we extend the model to allow h to take three values: $\{-1, 0, 1\}$, where $h = 0$ captures the case where the experimenter wants the subject to choose the natural action. We call the action following $h^T = 0$, $a^0(\zeta)$.

For simplicity we assume that the subject's prior belief is that each possibility is equally likely (i.e. is true with probability 1/3), so $E[h] = 0$. ϵ and η are also believed to take each value with probability 1/3 and are independent. $h^L \in \{-1, 0, 1\}$ and $h^T \in \{-1, 0, 1, \emptyset\}$ and p^L and p^T are defined as before. We maintain the assumption that the subject infers nothing when the experimenter does not send a demand treatment ($h^T = \emptyset$).

We show below that the beliefs can be written as:

$$E[h|h^L] = p^L h^L \tag{20}$$

$$E[h|h^T = \emptyset, h^L] = p^L h^L \tag{21}$$

$$E[h|h^T, h^L] = \frac{\frac{1}{3}(1-p^T)p^L h^L + \frac{1}{3}(1-p^L)p^T h^T + p^T p^L h^T \mathbb{I}[h^T = h^L]}{\frac{1}{3}\left(1 - p^T p^L\right) + p^T p^L \mathbb{I}[h^T = h^L]} \tag{22}$$

Bounding holds if $E[h|h^T = 1, h^L] \geq 0$ and $E[h|h^T = -1, h^L] \leq 0$. It is straightforward to check that the condition is the same as before: $p^T \geq p^L$.

What purpose, then, do $h^T = 0$ treatments serve? It is natural to think that demanding participants to take the natural action will eliminate demand effects, but under our assumptions, $h^T = 0$ does not in general elicit the natural action. Instead latent demand still influences the subject's action. We have:

$$E[h|h^T = 0, h^L] = \frac{\frac{1}{3}(1-p^T)p^L h^L}{\frac{1}{3}\left(1 - p^T p^L\right) + p^T p^L \mathbb{I}[h^L = 0]}$$

This expression equals zero if $p^T = 1$ (the demand treatment is perfectly informative), or $p^L h^L = 0$ (no latent demand), otherwise it has the same sign as $p^L h^L$. One interpretation is that while the subject takes at face value the experimenter's demand to choose the natural action, she might be unaware of the influence of other design features that nudge her in one direction or another.

Despite this negative result, $h^T = 0$ treatments can still be useful. First, they are informative about the *sign* of the bias due to latent demand. This is because $E[h|h^T = 0, h^L] \in [\min\{E[h|h^L], 0\}, \max\{E[h|h^L], 0\}]$ and therefore $a^0(\zeta) \in [\min\{a^L(\zeta), a(\zeta)\}, \max\{a^L(\zeta), a(\zeta)\}]$.[44] The action taken when $h^T = 0$ lies between the natural action and the action induced by latent demand, because the demand treatment shifts the subject's posterior toward zero.

Second, they can be used to obtain tighter bounds on $a(\zeta)$ if we know the direction of the latent demand effect. Suppose for example we know that $a^L(\zeta) \geq a(\zeta)$ (either from prior information or because we ran a treatment with $h^T = 0$ and verified that $a^0(\zeta) \leq a^L(\zeta)$). Then, the interval $[a^-(\zeta), a^0(\zeta)]$ gives a valid and tighter bound on $a(\zeta)$ than $[a^-(\zeta), a^+(\zeta)]$. Formally $a(\zeta) \in [a^-(\zeta), a^0(\zeta)] \subseteq [a^-(\zeta), a^+(\zeta)]$.[45]

Finally, there is one important case in which $h^T = 0$ perfectly recovers the natural action, i.e. $a^0(\zeta) = a(\zeta)$. Suppose that instead of assuming that the signals h^T and h^L contain independent shocks, the subject perceives that h^L is a noisy signal of h^T. Formally, he believes that with probability $p^L < 1$, $h^L = h^T$ and with probability $(1-p^L)$, $h^L = \epsilon$. Then, when h^T and h^L disagree, he knows that h^L is pure noise, when they agree h^L contains no more information than h^T. Hence, the subject disregards h^L after observing h^T and $E[h|h^T, h^L] = p^T h^T$. Then, sending $h^T = 0$ recovers the natural action: $E[h|h^T = 0, h^L] = 0, \forall h^L$. An advantage of our bounds is that they are valid whether or not h^T or h^L are perceived as independent, in other words they are conservative relative to the approach of simply measuring $a^0(\zeta)$.

C.6.1 Derivation of beliefs with ternary signals

Recall that now $h \in \{-1, 0, 1\}$, $h^L \in \{-1, 0, 1\}$ and $h^T \in \{-1, 0, 1, \emptyset\}$.

To avoid clutter we suppress dependence on ζ. After observing h^L, the subject forms a posterior $E[h|h^L] = Pr(h = 1|h^L) \times 1 + Pr(h =$

[44]To see this, note that $\left|E[h|h^L] - E[h|h^T = 0, h^L]\right| \geq 0$ and both have the same sign.

[45]We thank Liad Weiss for pointing this out to us.

$-1|h^L) \times (-1)$. We can write this as:

$$
\begin{aligned}
E[h|h^L = y] &= Pr(h = 1|h^L = y) - Pr(h = -1|h^L = y) \\
&= \frac{A}{B} \\
A &= Pr(h^L = y|h = 1)Pr(h = 1) - Pr(h^L = y|h = -1)Pr(h = -1) \\
B &= Pr(h^L = y|h = 1)Pr(h = 1) + Pr(h^L = y|h = 0)Pr(h = 0) \\
&\quad + Pr(h^L = y|h = -1)Pr(h = -1)
\end{aligned}
$$

Since $Pr(h = j|h^L = y) = \frac{1}{3}(1 - p^L) + p^L \mathbb{I}[y = j]$ and $Pr(h = j) = \frac{1}{3}$ we have

$$
\begin{aligned}
A &= \frac{1}{3}\left[\left(\frac{1}{3}(1 - p^L) + p^L \mathbb{I}[y = 1]\right) - \left(\frac{1}{3}(1 - p^L) + p^L \mathbb{I}[y = -1]\right)\right] \\
&= \frac{1}{3}p^L \left[\mathbb{I}[y = 1] - \mathbb{I}[y = -1]\right] = \frac{1}{3}p^L h^L \\
B &= \frac{1}{3}\left[\left(\frac{1}{3}(1 - p^L) + p^L \mathbb{I}[y = 1]\right) + \left(\frac{1}{3}(1 - p^L) + p^L \mathbb{I}[y = 0]\right)\right. \\
&\quad \left. + \left(\frac{1}{3}(1 - p^L) + p^L \mathbb{I}[y = -1]\right)\right] \\
&= \frac{1}{3}
\end{aligned}
$$

So

$$
E[h|h^L = y] = p^L h^L \tag{23}
$$

just as before. Turning to beliefs following the demand treatments, as before we assume that when $h^T = \emptyset$, $E[h|h^T, h^L] = E[h|h^L]$. We have:

$$
\begin{aligned}
E[h|h^T, h^L] &= Pr(h = 1|h^T, h^L) - Pr(h = -1|h^T, h^L) \\
&= \frac{A}{B} \\
A &= Pr(h^T = x|h = 1, h^L = y)Pr(h = 1|h^L = y) \\
&- Pr(h^T = x|h = -1, h^L = y)Pr(h = -1|h^L = y) \\
B &= Pr(h^T = x|h = 1, h^L = y)Pr(h = 1|h^L = y) \\
&+ Pr(h^T = x|h = 0, h^L = y)Pr(h = 0|h^L = y) \\
&+ Pr(h^T = x|h = -1, h^L = y)Pr(h = -1|h^L = y, h^L = y)
\end{aligned}
$$

Using

$$
\begin{aligned}
Pr(h^T = x|h = j, h^L = y) &= \frac{1}{3}(1 - p^T) + p^T \mathbb{I}[x = j] \\
Pr(h = j|h^L = y) &= \frac{1}{3}(1 - p^L) + p^L \mathbb{I}[y = j]
\end{aligned}
$$

we have:

$$
\begin{aligned}
A &= \left(\frac{1}{3}(1-p^T) + p^T \mathbb{I}[x=1]\right)\left(\frac{1}{3}(1-p^L) + p^L \mathbb{I}[y=1]\right) \\
&- \left(\frac{1}{3}(1-p^T) + p^T \mathbb{I}[x=-1]\right)\left(\frac{1}{3}(1-p^L) + p^L \mathbb{I}[y=-1]\right) \\
&= \frac{1}{3}(1-p^T)p^L \mathbb{I}[y=1] + \frac{1}{3}(1-p^L)p^T \mathbb{I}[x=1] + p^T p^L \mathbb{I}[x=1]\mathbb{I}[y=1] \\
&- \frac{1}{3}(1-p^T)p^L \mathbb{I}[y=-1] - \frac{1}{3}(1-p^L)p^T \mathbb{I}[x=-1] - p^T p^L \mathbb{I}[x=-1]\mathbb{I}[y=-1] \\
&= \frac{1}{3}(1-p^T)p^L h^L + \frac{1}{3}(1-p^L)p^T h^T + p^T p^L h^T \mathbb{I}[h^T = h^L] \\
B &= \left(\frac{1}{3}(1-p^T) + p^T \mathbb{I}[x=1]\right)\left(\frac{1}{3}(1-p^L) + p^L \mathbb{I}[y=1]\right) \\
&+ \left(\frac{1}{3}(1-p^T) + p^T \mathbb{I}[x=0]\right)\left(\frac{1}{3}(1-p^L) + p^L \mathbb{I}[y=0]\right) \\
&+ \left(\frac{1}{3}(1-p^T) + p^T \mathbb{I}[x=-1]\right)\left(\frac{1}{3}(1-p^L) + p^L \mathbb{I}[y=-1]\right) \\
&= \frac{1}{3}(1-p^T)(1-p^L) + \frac{1}{3}p^T(1-p^L)\left(\mathbb{I}[x=1] + \mathbb{I}[x=0] + \mathbb{I}[x=-1]\right) \\
&+ \frac{1}{3}p^L(1-p^T)\left(\mathbb{I}[y=1] + \mathbb{I}[y=0] + \mathbb{I}[y=-1]\right) \\
&+ p^T p^L\left(\mathbb{I}[x=1]\mathbb{I}[y=1] + \mathbb{I}[x=0]\mathbb{I}[y=0] + \mathbb{I}[x=-1]\mathbb{I}[y=-1]\right) \\
&= \frac{1}{3}\left(1 - p^T p^L\right) + p^T p^L \mathbb{I}[h^T = h^L]
\end{aligned}
$$

So

$$
E[h|h^T, h^L] = \frac{\frac{1}{3}(1-p^T)p^L h^L + \frac{1}{3}(1-p^L)p^T h^T + p^T p^L h^T \mathbb{I}[h^T = h^L]}{\frac{1}{3}\left(1 - p^T p^L\right) + p^T p^L \mathbb{I}[h^T = h^L]} \tag{24}
$$

C.7 Computing confidence intervals

C.7.1 Confidence intervals for actions

Imbens and Manski (2004) show that asymptotically the probability that the estimate for the upper (lower) bound is lower (higher) than the true value can be ignored when making inference. Thus, one can construct one-sided intervals with confidence level α around both the upper and the lower bound. The 95 percent confidence interval for the true demand-free behavior is thus given by:

$$CI() = [a^-(\zeta) - \overline{C_N}\frac{\widehat{\sigma^-}}{\sqrt{N}},\, a^+(\zeta) + \overline{C_N}\frac{\widehat{\sigma^+}}{\sqrt{N}}]$$

Here, $\widehat{\sigma^-} = \sqrt{\widehat{Var(a^-(\zeta))}}$ and $\widehat{\sigma^+} = \sqrt{\widehat{Var(a^+(\zeta))}}$, and $\overline{C_N}$ satisfies

$$\Phi\left(\overline{C_N} + \sqrt{N}\frac{a^+(\zeta) - a^-(\zeta)}{\max(\widehat{\sigma^-},\, \widehat{\sigma^+})}\right) - \Phi(-\overline{C_N}) = 0.90.$$

The 95 percent confidence interval for the set $[a^-(\zeta), a^+(\zeta)]$ is given by:

$$CI() = [a^-(\zeta) - \overline{C_N}\frac{\widehat{\sigma^-}}{\sqrt{N}},\, a^+(\zeta) + \overline{C_N}\frac{\widehat{\sigma^+}}{\sqrt{N}}],$$

where $\overline{C_N}$ satisfies

$$\Phi\left(\overline{C_N} + \sqrt{N}\frac{a^+(\zeta) - a^-(\zeta)}{\max(\widehat{\sigma^-},\, \widehat{\sigma^+})}\right) - \Phi(-\overline{C_N}) = 0.95.$$

C.7.2 Confidence intervals for treatment effects

We also outline how one can compute confidence intervals for the treatment effects $[a(\zeta_1) - a(\zeta_0)]$ and for the set defined by the upper and lower bounds for treatment effects as given by our demand treatments:[46] $[a(\zeta_1) - a(\zeta_0)] \in [a^-(\zeta_1) - a^+(\zeta_0), a^+(\zeta_1) - a^-(\zeta_0)]$

For simplicity we denote the lower bound, $[a^-(\zeta_1) - a^+(\zeta_0)]$, as T^- and

[46]We provide a Stata package, `demandbounds`, which computes the upper and lower bound of confidence intervals for mean behavior and treatment effects.

the upper bound, $[a^+(\zeta_1) - a^-(\zeta_0)]$, as T^+. The 95 percent confidence interval for the true demand-free treatment effect is given by:

$$CI() = [T^- - \overline{C_N}\frac{\widehat{\sigma^{T-}}}{\sqrt{N}}, T^+ + \overline{C_N}\frac{\widehat{\sigma^{T+}}}{\sqrt{N}}].$$

Here, $\widehat{\sigma^{T-}} = \sqrt{\widehat{Var(T^-)}}$ and $\widehat{\sigma^{T+}} = \sqrt{\widehat{Var(T^+)}}$, and $\overline{C_N}$ satisfies

$$\Phi\left(\overline{C_N} + \sqrt{N}\frac{T^+ - T^-}{\max(\widehat{\sigma^{T-}}, \widehat{\sigma^{T+}})}\right) - \Phi(-\overline{C_N}) = 0.90.$$

The 95 percent confidence interval for the set $[a^-(\zeta_1) - a^+(\zeta_0), a^+(\zeta_1) - a^-(\zeta_0)]$ is as follows:

$$CI() = [T^- - \overline{C_N}\frac{\widehat{\sigma^{T-}}}{\sqrt{N}}, T^+ + \overline{C_N}\frac{\widehat{\sigma^{T+}}}{\sqrt{N}}],$$

where

$$\Phi\left(\overline{C_N} + \sqrt{N}\frac{T^+(\tau) - T^-(\tau)}{\max(\widehat{\sigma^{T-}}, \widehat{\sigma^{T+}}}\right) - \Phi(-\overline{C_N}) = 0.95.$$

D Structural estimation appendix

This section outlines step by step how the parameters are constructed in our NLLS estimation of the structural model in section **4.3.**

D.1 Data and parameter adjustments

First, we follow DP exactly in rounding effort scores to the nearest 100 (except for those in range $[1, 49]$ which we round to 25). This is because incentives were paid per 100 points, and we wish to avoid modeling effort choices that lie between two 100 point thresholds. We refer the reader to DP for further details.

Second, we make a couple of adjustments pre and post-estimation. First, we divide the rounded scores by 100. In other words, if effort a is measured in points, we compute $a' = a/100$ which is measured in hundreds of points. Second, we multiply the incentive, ζ, which is measured in cents per point, by 100 to express it as $\zeta' = 100\zeta$ which is measured in cents per 100 points. These transformations were helpful in achieving convergence of the estimator, which otherwise occasionally suffered from underflow problems. However they change the interpretation of the parameters. Specifically, the intrinsic motivation parameter s and the preference for pleasing the experimenter, ϕ, will both be measured in units equivalent to cents per 100 points, while the cost function parameters will be expressed for effort measured in hundreds of points.

To aid comparability with DP we therefore re-transform the parameters after estimation. DP present their estimates of incentive parameters (which in our case are s and ϕ) in the same units, cents per 100 points, so we do not need to correct them. k and γ are reported for effort measured in points, so we transform our estimates for comparability. We derive the adjustments as follows. First, for the power cost function, we have:

$$U = (s + \zeta + \phi E[h|h^T, h^L])a - \frac{ka^{1+\gamma}}{1+\gamma}$$

Let $a' = \frac{a}{100}$ and $\zeta' = 100\zeta$. Then:

$$\begin{aligned} U &= \left(s + \frac{\zeta'}{100} + \phi E[h|h^T, h^L]\right) 100a' - \frac{k(100a')^{1+\gamma}}{1+\gamma} \\ &= \left(100s + \zeta' + 100\phi E[h|h^T, h^L]\right) a' - \frac{k(100a')^{1+\gamma}}{1+\gamma} \end{aligned}$$

giving rise to first-order condition:

$$\begin{aligned} 0 &= \left(100s + \zeta' + 100\phi E[h|h^T, h^L]\right) - ka'^{\gamma}100^{1+\gamma} \\ a' &= \left(\frac{100s + \zeta' + 100\phi E[h|h^T, h^L]}{k100^{1+\gamma}}\right)^{\frac{1}{\gamma}} \\ \log(a') &= \frac{1}{\gamma} \log\left(\frac{s^* + \zeta' + \phi^* E[h|h^T, h^L]}{k^*}\right) \end{aligned}$$

where $s^* = 100s$, $\phi^* = 100\phi$ and $k^* = 100^{1+\gamma}k$. We leave s^* and ϕ^*, (which are in equivalent units to cents per 100 points) untransformed for comparability with DP. In the tables we report $k = k^*/100^{1+\gamma}$ and its standard error, computed via the delta method.

For the exponential cost function we have:

$$\begin{aligned} U &= (s + \zeta + \phi E[h|h^T, h^L])a - \frac{k}{\gamma}\exp(\gamma a) \\ &= (s^* + \zeta' + \phi^* E[h|h^T, h^L])a' - \frac{k}{\gamma}\exp(100\gamma a') \end{aligned}$$

implying first-order condition:

$$
\begin{aligned}
0 &= s^* + \zeta' + \phi^* E[h|h^T, h^L] - 100k\exp(100\gamma a') \\
a' &= \frac{1}{100\gamma}\log\left(\frac{s^* + \zeta' + \phi^* E[h|h^T, h^L]}{100k}\right) \\
&= \frac{1}{\gamma^*}\log\left(\frac{s^* + \zeta' + \phi^* E[h|h^T, h^L]}{k^*}\right)
\end{aligned}
$$

where $s^* = 100s$, and $\phi^* = 100\phi$ as before, while $\gamma^* = 100\gamma$, $k^* = 100k$. In the tables we report $\gamma = \gamma^*/100$ and $k = k^*/100$.

D.2 Error term

To allow for the observed heterogeneity in effort, we follow DP in assuming heterogeneous effort costs, as follows. Let the cost of effort under power utility equal $ka^{1+\gamma}(1+\gamma)^{-1}\exp(-\gamma\epsilon)$ where $\epsilon \sim N(0, \sigma_\epsilon^2)$. Then our FOC becomes

$$
\begin{aligned}
0 &= \left(100s + \zeta' + 100\phi E[h|h^T, h^L]\right) - ka'^{\gamma} 100^{1+\gamma}\exp(-\gamma\epsilon) \\
a' &= \left(\frac{100s + \zeta' + 100\phi E[h|h^T, h^L]}{k100^{1+\gamma}}\right)^{\frac{1}{\gamma}}\exp(\epsilon) \\
\log(a') &= \frac{1}{\gamma}\log\left(\frac{100s + \zeta' + 100\phi E[h|h^T, h^L]}{k100^{1+\gamma}}\right) + \epsilon
\end{aligned}
$$

where ϵ becomes the error term in our NLLS routine. For the exponential cost, we follow DP and assume effort cost is $k\gamma^{-1}\exp(\gamma a)\exp(-\gamma\epsilon)$. Then our FOC becomes

$$
\begin{aligned}
0 &= s^* + \zeta' + \phi^* E[h|h^T, h^L] - 100k\exp(100\gamma a')\exp(-\gamma\epsilon) \\
a' &= \frac{1}{100\gamma}\log\left(\frac{s^* + \zeta' + \phi^* E[h|h^T, h^L]}{100k}\right) + \frac{\epsilon}{100} \\
&= \frac{1}{\gamma^*}\log\left(\frac{s^* + \zeta' + \phi^* E[h|h^T, h^L]}{k^*}\right) + \epsilon^*
\end{aligned}
$$

where $\epsilon^* = \epsilon/100$ forms the error term in our estimation.

D.3 Estimating equation

Finally, in our estimation we sometimes need to estimate the product $\phi^* E[h|h^L]$. We estimate this product directly, then transform by dividing by ϕ^*. Specifically, we estimate the following:

$$
\begin{aligned}
y_i = &\frac{1}{\beta_0} \log [\zeta_i' + \beta_1 + \beta_2(\text{pos\_demand}_i - \text{neg\_demand}_i) \\
&+ \beta_3 \times \text{no\_demand}_i \times \text{incentive\_0c}_i + \beta_4 \times \text{no\_demand}_i \times \text{incentive\_1c}_i \\
&+ \beta_5 \times \text{no\_demand}_i \times \text{incentive\_4c}_i] - \frac{1}{\beta_0} \log(\beta_6) + \varepsilon_i
\end{aligned}
$$

where $y = \log(a')$ or a' respectively, pos_demand, neg_demand and no_demand are dummies for our positive, negative and no demand treatments, while incentive_Xc is a dummy for the treatment with X cents per 100 points. Parameters are as follows: $\beta_0 = \gamma$ or γ^* respectively, $\beta_1 = s^*$, $\beta_2 = \phi^*$, $\beta_3 = \phi^* E[h|h^L(\zeta = 0)]$, $\beta_4 = \phi^* E[h|h^L(\zeta = 1)]$, $\beta_5 = \phi^* E[h|h^L(\zeta = 4)]$ and $\beta_6 = k^*$. We then compute the three values for $E[h|h^L]$ by dividing by β_2, i.e. β_3/β_2, β_4/β_2 and β_5/β_2. γ and k are computed by the transformations outlined above. Standard errors are computed by the delta method. In the specification where we restrict latent demand to be equal for the 1 cent and 4 cent treatments we impose $\beta_4 = \beta_5$.

D.4 Predicted values

One use of our structural estimates is to compute predicted effort when latent demand is shut down, i.e. when $\phi E[h|h^L] = 0$. To do this we need to make one more adjustment, namely to express intrinsic motivation in units of cents per point by dividing the estimates of s^* by 100, and to express ζ in cents per point (i.e. 0, 0.01 or 0.04 respectively). So, in terms of our estimated parameters, predicted effort (or log effort in the power cost case) is:

$$\frac{1}{\gamma}\log\left(\frac{s+\zeta}{k}\right)$$

D.5 Comparison with DP

Our parameter estimates are quite different from DP's, so we briefly explore why. DP (Figure 2) provide a graphical representation of their estimates in terms of marginal cost and marginal benefit of effort, which we can replicate here to compare our estimates. We focus on the exponential cost case, comparing our specification (4) (which assumes no latent demand and uses only the no demand treatment groups) with theirs from Table 5, panel A specification (4).[47]

Figure A.14 plots, for our estimates and theirs, the marginal cost function, minus intrinsic motivation: $c'(a) - s$. By the first-order condition, optimal effort is the point at which this function equals ζ, which takes values in $\{0, 0.01, 0.04, 0.1\}$. We also plot mean effort under each no demand treatment in our experiment and in DPs. It is immediately clear that the differences in the parameter estimates are driven by lower effort under the 0c and 1c treatments in our experiment than in DPs.[48]

[47]DP's Figure 2 shows the comparison between predicted and observed mean effort, and is computed from their minimum distance estimation (MDE) parameters. Since we focus on NLLS estimation, the figure for the power cost is not very informative, because the estimation matches mean log effort (MDE matches mean effort). We therefore focus on the exponential case.

[48]By construction our estimated function exactly equals mean effort at the treatment values. DP's marginal cost function does not pass through the 4 cent point because it was estimated from the 0, 1 and 10 cent treatments with the 4 cent treatment included for out-of-sample evaluation. Other small differences due to rounding in their reported parameters.

Figure A.13: Marginal cost and benefit of effort (exponential), comparison with DellaVigna and Pope (2016)

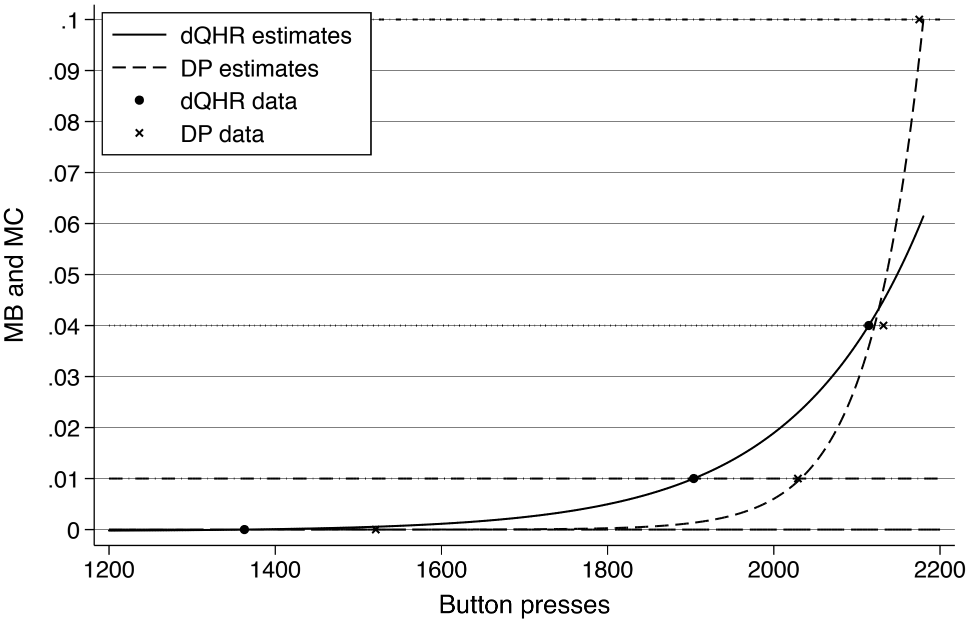


Figure A.14: Marginal cost and benefit of effort (exponential), comparison with DellaVigna and Pope (2016)

E Online Appendix: Pre-specified Tables

E.1 Pre-analysis Plan 1

Table A.19: Strong Demand (Experiment 1)

	(1)	(2)	(3)	(4)	(5)
Positive Demand	0.240***	0.196***	0.278***	0.016	0.453***
	(0.035)	(0.049)	(0.048)	(0.190)	(0.058)
Negative Demand	-0.248***	-0.257***	-0.269***	-0.208	-0.203***
	(0.036)	(0.051)	(0.048)	(0.176)	(0.055)
Positive demand × interactant		0.089	-0.073	0.235	
		(0.070)	(0.070)	(0.193)	
Negative demand × interactant		0.019	0.041	-0.043	
		(0.072)	(0.072)	(0.179)	
Interactant		-0.091	-0.044	-0.083	
		(0.051)	(0.051)	(0.141)	
Positive Demand × Risk					-0.255**
					(0.084)
Negative Demand × Risk					-0.033
					(0.083)
Positive Demand × Time					-0.392***
					(0.085)
Negative Demand × Time					-0.116
					(0.087)
Constant	-0.145***	-0.100**	-0.122***	-0.065	-0.335***
	(0.025)	(0.035)	(0.034)	(0.139)	(0.040)
Interactant		Monetary Incentive	Male	Attention	
Adjusted R^2	0.040	0.041	0.041	0.040	0.051
Positive demand ≤ 0	0.000	0.000	0.000	0.466	0.000
Adjusted p-value	0.010	0.001	0.001	0.307	0.001
Negative demand ≥ 0	0.000	0.000	0.000	0.118	0.000
Adjusted p-value	0.010	0.001	0.001	0.307	0.001
Positive demand = negative demand	0.000	0.000	0.000	0.182	0.000
Adjusted p-value	0.010	0.001	0.001	0.307	0.001
(Positive demand - negative demand)* interaction = 0		0.319	0.105	0.105	
Adjusted p-value		0.086	0.027	0.307	
Risk*(pos - neg) = Time*(pos - neg)					0.533
Adjusted p-value					0.179
Risk*(positive demand - negative demand) = 0					0.007
Adjusted p-value					0.005
Time*(positive demand - negative demand) = 0					0.001
Adjusted p-value					0.001
Joint F-test					.001
Observations	4495	4495	4495	4495	4495

Notes: This table summarizes the results from experiment 1. The outcome variable is normalized at the game-level using the mean and standard deviation of the negative demand group. Robust standard errors are in parentheses. False-discovery rate adjusted p-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.20: Beliefs about the experimental objective and hypothesis: Strong Demand

	Belief: Want High	Belief: Expect High
Positive - Negative	0.275***	0.180***
	(0.017)	(0.018)
Adjusted p-value	[0.001]	[0.001]
Positive - Neutral	0.160***	0.143***
	(0.017)	(0.018)
Adjusted p-value	[0.001]	[0.001]
Negative - Neutral	-0.115***	-0.037**
	(0.018)	(0.018)
Adjusted p-value	[0.001]	[0.007]
Mean (No Demand)	0.542	0.450
Observations	4495	4495

Notes: The outcome variables take value one if the respondents believed that the experimenter wanted (column 1) or expected (column 2) a high action. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

E.2 Pre-analysis Plan 2

Table A.21: Weak Demand (Experiment 2)

	(1)	(2)	(3)	(4)	(5)
Positive Demand	0.126**	0.151*	0.083	0.067	0.203***
	(0.043)	(0.060)	(0.056)	(0.116)	(0.054)
Negative Demand	-0.040	0.031	-0.035	-0.023	-0.042
	(0.042)	(0.060)	(0.055)	(0.109)	(0.054)
Pos. demand × interactant		-0.053	0.091	0.069	-0.149
		(0.085)	(0.086)	(0.124)	(0.085)
Neg. demand × interactant		-0.142	-0.011	-0.019	0.007
		(0.083)	(0.084)	(0.118)	(0.083)
Interactant		-0.063	-0.029	-0.218**	0.193**
		(0.060)	(0.061)	(0.081)	(0.060)
Interactant		Monetary Incentive	Male	Attention	Risk
Adjusted R-squared	0.005	0.010	0.005	0.009	0.012
Pos. demand ≤ 0	0.002	0.006	0.068	0.280	0.000
Adjusted p-value	0.010	0.020	0.150	0.970	0.010
Neg. demand ≥ 0	0.168	0.700	0.264	0.415	0.221
Adjusted p-value	0.050	0.530	0.150	0.970	0.080
Pos. demand = neg. demand	0.000	0.043	0.034	0.448	0.000
Adjusted p-value	0.010	0.060	0.150	0.970	0.010
(Pos. - neg.) × interactant = 0		0.283	0.222	0.493	0.062
Adjusted p-value		0.230	0.150	0.970	0.040
Observations	2964	2964	2964	2964	2964

Notes: This table summarizes the results from experiment 2. The outcome variable is normalized at the game-level using the mean and standard deviation of the negative demand group. Robust standard errors are in parentheses. False-discovery rate adjusted p-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.22: Difference in response to demand between experiment 1 and experiment 2

	(1)	(2)	(3)
Positive Demand=1	0.126**	0.203***	0.054
	(0.043)	(0.054)	(0.065)
Experiment 1=1	-0.135**	-0.140*	-0.125
	(0.043)	(0.056)	(0.064)
Positive Demand=1 × Experiment 1=1	0.202***	0.251**	0.145
	(0.060)	(0.079)	(0.090)
Negative Demand=1	-0.040	-0.042	-0.035
	(0.042)	(0.054)	(0.063)
Negative Demand=1 × Experiment 1=1	-0.175**	-0.161*	-0.202*
	(0.059)	(0.077)	(0.088)
Constant	-0.097**	-0.195***	-0.003
	(0.030)	(0.039)	(0.046)
Sample	All	Dictator Game	Investment
Adjusted R^2	0.034	0.056	0.020
H_0 : (Positive Demand - Negative Demand)*Interaction = 0	0.000	0.000	0.000
Adjusted p-value	0.001	0.001	0.001
Observations	5971	2990	2981

Notes: This table uses data from the investment game and dictator game in experiments 1 and 2. The dummy experiment 1 takes value 1 for respondents from experiment 1. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.23: Beliefs about the experimental objective and hypothesis: Weak Demand (Experiment 2)

	Belief: Want High	Belief: Expect High
Positive - Negative	0.334***	0.403***
	(0.021)	(0.020)
Adjusted p-value	[0.001]	[0.001]
Positive - Neutral	0.172***	0.217***
	(0.022)	(0.022)
Adjusted p-value	[0.001]	[0.001]
Negative - Neutral	-0.162***	-0.186***
	(0.022)	(0.020)
Adjusted p-value	[0.001]	[0.001]
Mean (No Demand)	0.485	0.392
Observations	2964	2964

Notes: This table uses data from all respondents who completed experiment 2. The outcome variables take value one if the respondents believed that the experimenter wanted (column 1) or expected (column 2) a high action. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.24: Beliefs about whether the experiment is incentivized

	(1) Belief: Real Money
Monetary Incentive	0.368***
	(0.016)
Control Mean	0.138
R^2	0.154
Observations	2964

Notes: This table uses data from all respondents who completed experiment 2. The outcome variable takes value one if the respondent believes that the tasks in the experiment involve real money and value zero otherwise. Notes go here. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.25: Attrition

	(1) Finished
Positive Demand	0.00285
	(0.004)
Negative Demand	0.00115
	(0.004)
Mean (no demand)	0.988
R^2	0.000141
Observations	2993

Notes: This table uses data from all respondents who started experiment 2. The outcome variable takes value one if the respondent completed the experiment. Finished takes value one for all respondents who completed the experiment. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

E.3 Pre-analysis Plan 3

Table A.26: Effort (z-scored) with strong demand

	(1)	(2)	(3)
Positive Demand	0.206***	0.328***	0.319**
	(0.061)	(0.085)	(0.107)
Negative Demand	-0.309***	-0.447***	-0.197
	(0.061)	(0.084)	(0.103)
Positive demand × interactant		-0.243*	-0.197
		(0.123)	(0.129)
Negative demand × interactant		0.288*	-0.200
		(0.121)	(0.126)
Interactant		0.091	0.143
		(0.088)	(0.093)
Constant	0.069	0.023	-0.012
	(0.044)	(0.061)	(0.078)
Interactant		1-cent incentive	Male
Adjusted R^2	0.047	0.060	0.047
Positive demand ≤ 0	0.000	0.000	0.001
Adjusted p-value	0.010	0.001	0.002
Negative demand ≥ 0	0.000	0.000	0.028
Adjusted p-value	0.010	0.001	0.018
Positive demand = negative demand	0.000	0.000	0.000
Adjusted p-value	0.010	0.001	0.001
(Positive demand - negative demand)* interaction = 0		0.000	0.975
Adjusted p-value		0.001	0.322
Observations	1452	1452	1452

Notes: This table summarizes the results from experiment 3. The outcome variable is normalized at the game level using the mean and standard deviation of the negative demand group. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.27: Beliefs: Effort with strong demand

	Belief: Want High	Belief: Expect High
Positive - Negative	0.459***	0.416***
	(0.027)	(0.028)
Adjusted p-value	[0.001]	[0.001]
Positive - Neutral	0.168***	0.192***
	(0.026)	(0.028)
Adjusted p-value	[0.001]	[0.001]
Negative - Neutral	-0.291***	-0.224***
	(0.031)	(0.031)
Adjusted p-value	[0.001]	[0.001]
Mean (No Demand)	0.689	0.639
Observations	1452	1452

Notes: The outcome variables take value one if the respondents believed that the experimenter wanted (column 1) or expected (column 2) a high action. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.28: Attrition

	(1) Finished
Positive Demand	0.000252
	(0.010)
Negative Demand	0.00353
	(0.010)
Mean (no demand)	0.988
R^2	0.0000802
Observations	1753

Notes: The outcome variable takes value one if the respondent completed the experiment. Finished takes value one for all respondents who completed the experiment. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

E.4 Pre-analysis Plan 4

Table A.29: Demand: Representative Sample with strong and weak demand treatments (Experiment 4)

	(1)	(2)	(3)	(4)	(5)
Positive Demand	0.284***	0.196**	0.325***	0.247***	0.554***
	(0.055)	(0.063)	(0.063)	(0.061)	(0.064)
Negative Demand	-0.157**	-0.034	-0.221***	-0.082	-0.034
	(0.055)	(0.064)	(0.061)	(0.060)	(0.064)
Pos. demand × interactant		0.175**	-0.084	0.112	-0.538***
		(0.064)	(0.064)	(0.064)	(0.062)
Neg. demand × interactant		-0.238***	0.136*	-0.219***	-0.251***
		(0.063)	(0.064)	(0.063)	(0.063)
Interactant		Strong Demand	Male	Attention	Risk
Adjusted R-squared	0.031	0.038	0.033	0.035	0.060
Pos. demand ≤ 0	0.000	0.001	0.000	0.000	0.000
Adjusted p-value	0.010	0.010	0.010	0.010	0.010
Neg. demand ≥ 0	0.002	0.297	0.000	0.086	0.297
Adjusted p-value	0.010	0.080	0.010	0.020	0.080
Pos. = neg. demand	0.000	0.000	0.000	0.000	0.000
Adjusted p-value	0.010	0.010	0.010	0.010	0.010
(Pos. - neg.) × interactant = 0		0.000	0.015	0.000	0.001
Adjusted p-value		0.010	0.010	0.010	0.010
Observations	2941	2941	2941	2941	2941

Notes: This table summarizes the results from experiment 4. The outcome variable is normalized at the game level using the mean and standard deviation of the negative demand group. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.30: Demand Sensitivty by game: Representative vs. MTurk Sample

	(1)	(2)	(3)
Positive Demand=1	0.256***	0.419***	0.091
	(0.043)	(0.057)	(0.064)
Representative Sample=1	0.517***	0.848***	0.201**
	(0.054)	(0.075)	(0.075)
Positive Demand=1 × Representative Sample=1	0.028	-0.076	0.124
	(0.070)	(0.097)	(0.097)
Negative Demand=1	-0.153***	-0.046	-0.256***
	(0.042)	(0.056)	(0.061)
Negative Demand=1 × Representative Sample=1	-0.004	-0.200*	0.169
	(0.069)	(0.096)	(0.096)
Constant	-0.216***	-0.335***	-0.099*
	(0.031)	(0.041)	(0.045)
Sample	All	Dictator Game	Investment
Adjusted R^2	0.093	0.165	0.041
H_0 : (Positive Demand - Negative Demand)*Repres. Sample = 0	0.593	0.149	0.597
Adjusted p-value	0.805	0.805	0.805
Observations	5948	3004	2944

Notes: This table uses data from the incentivized MTurk respondents from experiments 1 and 2 and the representative online panel (experiment 4). Representative Sample is a dummy variable taking value 1 for respondents from the representative online panel and value zero for the MTurk respondents. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.31: Beliefs about the experimental objective and hypothesis: Represenrative Sample

	Belief: Want High	Belief: Expect High
Positive - Negative	0.207***	0.205***
	(0.020)	(0.020)
Adjusted p-value	[0.001]	[0.001]
Positive - Neutral	0.068***	0.092***
	(0.024)	(0.025)
Adjusted p-value	[0.001]	[0.001]
Negative - Neutral	-0.139***	-0.114***
	(0.025)	(0.025)
Adjusted p-value	[0.001]	[0.001]
Mean (No Demand)	0.601	0.510
Observations	2939	2941

Notes: The outcome variables take value one if the respondents believed that the experimenter wanted (column 1) or expected (column 2) a high action. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.32: Attrition

	(1) Finished
Positive Demand	0.00148
	(0.005)
Negative Demand	-0.00223
	(0.005)
Mean (no demand)	0.988
R^2	0.000329
Observations	2966

Notes: The outcome variable takes value one if the respondent completed the experiment. Finished takes value one for all respondents who completed the experiment. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

E.5 Pre-analysis Plan 5

Table A.33: Differences in response to demand across games

	(1)	(2)
Sensitivity=1	1.024***	0.288*
	(0.127)	(0.120)
Ambiguity	0.129	0.018
	(0.107)	(0.100)
DG	0.079	0.037
	(0.090)	(0.078)
Effort: incentive	0.316**	0.104
	(0.101)	(0.098)
Effort: no incentive	-0.063	0.067
	(0.102)	(0.099)
Lying	0.240*	0.040
	(0.121)	(0.100)
Risk	0.108	0.042
	(0.090)	(0.078)
Time	0.076	0.036
	(0.098)	(0.097)
Trust	0.126	0.033
	(0.106)	(0.104)
UG 1	0.129	0.032
	(0.115)	(0.102)
UG 2	0.233*	0.041
	(0.115)	(0.098)
Sensitivity=1 × Ambiguity	-0.565***	-0.110
	(0.160)	(0.157)
Sensitivity=1 × DG	-0.309*	-0.022
	(0.138)	(0.130)
Sensitivity=1 × Effort: incentive	-0.781***	-0.204
	(0.153)	(0.153)
Sensitivity=1 × Effort: no incentive	-0.250	-0.357*
	(0.152)	(0.157)
Sensitivity=1 × Lying	-0.427*	-0.248
	(0.173)	(0.157)
Sensitivity=1 × Risk	-0.593***	-0.157
	(0.136)	(0.131)
Sensitivity=1 × Time	-0.644***	-0.288
	(0.144)	(0.154)
Sensitivity=1 × Trust	-0.470**	-0.209
	(0.160)	(0.159)
Sensitivity=1 × UG 1	-0.338*	-0.113
	(0.168)	(0.165)
Sensitivity=1 × UG 2	-0.277	-0.014
	(0.168)	(0.157)
Constant	-0.361***	-0.033
	(0.083)	(0.070)
Treatment	Strong	Weak
Adjusted R^2	0.084	0.007
P-value(Omnibus F-Test)	0.000	0.063
Adjusted p-values	0.001	0.043
P-value(Omnibus F-Test): without effort tasks	0.000	0.166
Adjusted p-values	0.001	0.090
Observations	7523	6599

Notes: We pool all observations across all experiments. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.34: Differences in response to strong vs. weak demand treatments

	(1) Z-scored behavior
Strong Demand × Sensitivity	0.421***
	(0.035)
Sensitivity	0.153***
	(0.025)
R^2	0.0429
Observations	14122

Notes: We pool all observations across all experiments. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

F Online Appendix: Balance Tables and Summary statistics

Table A.35: Balance Table: Experiment 1 (Strong Demand)

	No demand	Pos. demand	Neg. demand	P-value(Pos. demand - no demand)	P-value(Neg. demand - no demand)	P-value(Pos. demand - neg. demand)	Observations
Male	0.511	0.520	0.497	0.641	0.432	0.209	4495
Income	51545.455	52414.421	53387.833	0.402	0.072	0.344	4008
Age	36.195	36.434	36.382	0.557	0.655	0.898	4495
Household Size	3.714	3.649	3.625	0.205	0.087	0.639	4495
White	0.773	0.785	0.773	0.444	0.983	0.434	4495
Black	0.070	0.066	0.072	0.669	0.867	0.553	4495
Hispanic	0.053	0.057	0.055	0.597	0.821	0.766	4495
Asian	0.080	0.064	0.076	0.105	0.709	0.216	4495
Full-time employment	0.484	0.507	0.521	0.208	0.049	0.464	4495
Part-time employment	0.127	0.121	0.114	0.607	0.283	0.569	4495
Unemployed	0.143	0.133	0.129	0.402	0.272	0.785	4495
Bachelor Degree	0.353	0.371	0.389	0.300	0.043	0.313	4495
Conservative	0.232	0.238	0.241	0.689	0.535	0.822	4457
Number of HITs	9366.555	9202.861	8642.955	0.777	0.212	0.324	4495
Joint							

Notes: In this table we present evidence on the experimental integrity in Experiment 1. The p-value of the joint-F-test when comparing covariates in the positive and negative demand condition is 0.9091. The p-value of the joint-F-test when comparing covariates in the positive and no-demand demand condition is 0.7123. The p-value of the joint-F-test when comparing covariates in the negative and no-demand demand condition is 0.2543.

Table A.36: Balance Table: Experiment 2 (Weak Demand)

	No demand	Pos. demand	Neg. demand	P-value(Pos. demand - no demand)	P-value(Neg. demand - no demand)	P-value(Pos. demand - neg. demand)	Observations
Male	0.466	0.466	0.478	0.984	0.577	0.592	2964
Income	51010.333	51307.604	52093.679	0.815	0.384	0.526	2625
Age	35.897	35.856	35.168	0.935	0.142	0.166	2964
Household Size	3.696	3.688	3.761	0.900	0.314	0.258	2964
White	0.784	0.760	0.748	0.203	0.055	0.526	2964
Black	0.070	0.076	0.077	0.593	0.557	0.963	2964
Hispanic	0.054	0.051	0.057	0.827	0.760	0.600	2964
Asian	0.066	0.070	0.089	0.714	0.056	0.124	2964
Full-time employment	0.494	0.464	0.468	0.185	0.249	0.854	2964
Part-time employment	0.130	0.099	0.125	0.032	0.735	0.069	2964
Unemployed	0.101	0.140	0.127	0.009	0.065	0.417	2964
Bachelor Degree	0.367	0.353	0.377	0.503	0.642	0.256	2964
Conservative	0.273	0.253	0.243	0.328	0.128	0.594	2941
Number of HITs	5849.696	5629.887	5403.884	0.693	0.415	0.673	2964

Notes: In this table we present evidence on the experimental integrity in Experiment 2. The p-value of the joint-F-test when comparing covariates in the positive and negative demand condition is 0.6464. The p-value of the joint-F-test when comparing covariates in the positive and no-demand demand condition is 0.2297. The p-value of the joint-F-test when comparing covariates in the negative and no-demand demand condition is 0.4443.

Table A.37: Balance Table: Experiment 3 (Effort Experiment with strong demand)

	No demand	Pos. demand	Neg. demand	P-value(Pos. demand - no demand)	P-value(Neg. demand - no demand)	P-value(Pos. demand - neg. demand)	Observations
Male	0.557	0.576	0.535	0.521	0.463	0.209	1699
Income	33600.823	32204.082	32458.333	0.164	0.261	0.814	1699
Age	37.449	37.378	36.556	0.922	0.213	0.300	1699
Household Size	3.750	3.780	3.763	0.724	0.879	0.847	1699
White	0.752	0.784	0.760	0.193	0.730	0.389	1699
Black	0.110	0.084	0.083	0.127	0.124	0.985	1699
Hispanic	0.055	0.024	0.046	0.006	0.479	0.072	1699
Asian	0.064	0.071	0.075	0.638	0.485	0.831	1699
Full-time employment	0.508	0.496	0.540	0.691	0.275	0.174	1699
Part-time employment	0.125	0.127	0.106	0.930	0.320	0.325	1699
Unemployed	0.106	0.122	0.106	0.368	0.972	0.428	1699
Bachelor Degree	0.395	0.355	0.371	0.157	0.396	0.611	1699
Republican	0.251	0.288	0.271	0.158	0.445	0.557	1699

Notes: In this table we present evidence on the integrity of the randomization in Experiment 3. The p-value of the joint-F-test when comparing covariates in the positive and negative demand condition is 0.8777. The p-value of the joint-F-test when comparing covariates in the positive and no-demand demand condition is 0.0966. The p-value of the joint-F-test when comparing covariates in the negative and no-demand demand condition is 0.4331.

Table A.38: Balance Table: Experiment 4 (Representative Sample)

	No demand	Pos. demand	Neg. demand	P-value(Pos. demand - no demand)	P-value(Neg. demand - no demand)	P-value(Pos. demand - neg. demand)	Observations
Male	0.487	0.485	0.470	0.937	0.497	0.467	2941
Income	68432.773	65257.447	67142.857	0.233	0.632	0.393	2890
Age	47.923	46.922	47.853	0.226	0.933	0.168	2941
Household Size	3.335	3.311	3.335	0.694	0.998	0.648	2934
White	0.799	0.772	0.784	0.188	0.483	0.468	2935
Black	0.073	0.069	0.061	0.781	0.376	0.453	2935
Hispanic	0.051	0.064	0.061	0.262	0.373	0.794	2935
Asian	0.043	0.061	0.062	0.086	0.077	0.938	2935
Full-time employment	0.500	0.484	0.495	0.522	0.848	0.590	2941
Part-time employment	0.076	0.079	0.092	0.802	0.238	0.267	2941
Unemployed	0.067	0.050	0.052	0.136	0.191	0.822	2941
Bachelor Degree	0.329	0.352	0.329	0.326	1.000	0.238	2941
Conservative	0.350	0.351	0.351	0.958	0.980	0.974	2804

Notes: In this table we present evidence on the integrity of the randomization in Experiment 4. The p-value of the joint-F-test when comparing covariates in the positive and negative demand condition is 0.7723. The p-value of the joint-F-test when comparing covariates in the positive and no-demand demand condition is 0.4676. The p-value of the joint-F-test when comparing covariates in the negative and no-demand demand condition is 0.6403.

Table A.39: Balance Table: Experiment 5 (Many Task experiment)

	Pos. demand	Neg. demand	P-value(Pos. demand - neg. demand)	Observations
Male	0.452	0.473	0.135	5068
Income	53223.655	52705.464	0.507	4500
Age	37.314	37.181	0.685	5068
Household Size	3.710	3.651	0.149	5068
White	0.769	0.774	0.626	5068
Black	0.077	0.072	0.479	5068
Hispanic	0.048	0.049	0.978	5068
Asian	0.077	0.078	0.880	5068
Full-time employment	0.513	0.517	0.785	5068
Part-time employment	0.116	0.113	0.748	5068
Unemployed	0.125	0.140	0.129	5068
Bachelor Degree	0.376	0.371	0.764	5068
Conservative	0.262	0.256	0.642	5042
Number of HITs	9341.149	8553.308	0.069	5068

Notes: In this table we present evidence on the integrity of the randomization in Experiment 5. The p-value of the joint-F-test when comparing covariates in the positive and negative demand condition is 0.2164.

Table A.40: Balance Table: Experiment 6 (Effort Experiment with weak demand treatments)

	Pos. demand	Neg. demand	P-value(Pos. demand - neg. demand)	Observations
Male	0.545	0.557	0.748	775
Income	32235.142	32474.227	0.845	775
Age	37.323	37.668	0.685	775
Household Size	3.729	3.683	0.663	775
White	0.757	0.732	0.423	775
Black	0.083	0.082	0.991	775
Hispanic	0.054	0.072	0.306	775
Asian	0.080	0.075	0.780	775
Full-time employment	0.548	0.528	0.588	775
Part-time employment	0.129	0.093	0.107	775
Unemployed	0.127	0.124	0.903	775
Bachelor Degree	0.432	0.379	0.136	775
Conservative	0.264	0.325	0.066	770

Notes: In this table we present evidence on the integrity of the randomization in Experiment 6. The p-value of the joint-F-test when comparing covariates in the positive and negative demand condition is 0.2556.

Table A.41: Summary Statistics: Pooled across all experiments

	Mean	SD	Median	Min.	Max.	Obs.
Male	0.49	0.50	0.00	0.00	1.00	17942
Income	51960.66	33091.17	45000.00	5000.00	225000.00	16497
Age	38.44	13.10	35.00	17.00	116.00	17942
Household Size	3.63	1.40	3.00	2.00	13.00	17935
White	0.77	0.42	1.00	0.00	1.00	17936
Black	0.07	0.26	0.00	0.00	1.00	17936
Hispanic	0.05	0.22	0.00	0.00	1.00	17936
Asian	0.07	0.26	0.00	0.00	1.00	17936
Full-time employment	0.50	0.50	1.00	0.00	1.00	17942
Part-time employment	0.11	0.32	0.00	0.00	1.00	17942
Unemployed	0.12	0.32	0.00	0.00	1.00	17942
Bachelor Degree	0.37	0.48	0.00	0.00	1.00	17942
Conservative	0.27	0.44	0.00	0.00	1.00	16014
Number of HITs	8209.03	14913.36	2500.00	750.00	75000.00	12527

Notes: This Table summarizes the main covariates of all respondents across all 6 experiments.

Table A.42: Summary Statistics: Experiment 1 (Strong demand)

	Mean	SD	Median	Min.	Max.	Obs.
Male	0.51	0.50	1.00	0.00	1.00	4495
Income	52447.60	26624.13	55000.00	5000.00	100000.00	4008
Age	36.34	11.26	33.00	19.00	88.00	4495
Household Size	3.66	1.40	3.00	2.00	11.00	4495
White	0.78	0.42	1.00	0.00	1.00	4495
Black	0.07	0.25	0.00	0.00	1.00	4495
Hispanic	0.05	0.23	0.00	0.00	1.00	4495
Asian	0.07	0.26	0.00	0.00	1.00	4495
Full-time employment	0.50	0.50	1.00	0.00	1.00	4495
Part-time employment	0.12	0.33	0.00	0.00	1.00	4495
Unemployed	0.14	0.34	0.00	0.00	1.00	4495
Bachelor Degree	0.37	0.48	0.00	0.00	1.00	4495
Conservative	0.24	0.43	0.00	0.00	1.00	4457
Number of HITs	9075.19	15743.81	2500.00	750.00	75000.00	4495

Notes: This Table summarizes the main covariates of all respondents in experiment 1.

Table A.43: Summary Statistics: Experiment 2 (Weak demand)

	Mean	SD	Median	Min.	Max.	Obs.
Male	0.47	0.50	0.00	0.00	1.00	2964
Income	51474.29	26146.68	55000.00	5000.00	100000.00	2625
Age	35.64	11.08	33.00	19.00	81.00	2964
Household Size	3.72	1.43	3.00	2.00	13.00	2964
White	0.76	0.42	1.00	0.00	1.00	2964
Black	0.07	0.26	0.00	0.00	1.00	2964
Hispanic	0.05	0.23	0.00	0.00	1.00	2964
Asian	0.07	0.26	0.00	0.00	1.00	2964
Full-time employment	0.48	0.50	0.00	0.00	1.00	2964
Part-time employment	0.12	0.32	0.00	0.00	1.00	2964
Unemployed	0.12	0.33	0.00	0.00	1.00	2964
Bachelor Degree	0.37	0.48	0.00	0.00	1.00	2964
Conservative	0.26	0.44	0.00	0.00	1.00	2941
Number of HITs	5626.60	12144.69	1500.00	750.00	75000.00	2964

Notes: This Table summarizes the main covariates of all respondents in experiment 2.

Table A.44: Summary Statistics: Experiment 3 (Effort Experiment: Strong demand)

	Mean	SD	Median	Min.	Max.	Obs.
Male	0.56	0.50	1.00	0.00	1.00	1699
Income	32875.22	17276.75	35000.00	5000.00	85000.00	1699
Age	37.18	12.33	36.00	21.00	70.00	1699
Household Size	3.76	1.39	4.00	2.00	12.00	1699
White	0.76	0.43	1.00	0.00	1.00	1699
Black	0.09	0.29	0.00	0.00	1.00	1699
Hispanic	0.04	0.20	0.00	0.00	1.00	1699
Asian	0.07	0.25	0.00	0.00	1.00	1699
Full-time employment	0.51	0.50	1.00	0.00	1.00	1699
Part-time employment	0.12	0.33	0.00	0.00	1.00	1699
Unemployed	0.11	0.31	0.00	0.00	1.00	1699
Bachelor Degree	0.38	0.48	0.00	0.00	1.00	1699
Republican	0.27	0.44	0.00	0.00	1.00	1699

Notes: This Table summarizes the main covariates of all respondents in experiment 3.

Table A.45: Summary Statistics: Experiment 4 (Representative sample)

	Mean	SD	Median	Min.	Max.	Obs.
Male	0.48	0.50	0.00	0.00	1.00	2941
Income	66641.87	52841.78	62500.00	7500.00	225000.00	2890
Age	47.49	16.38	47.00	17.00	116.00	2941
Household Size	3.33	1.25	3.00	2.00	13.00	2934
White	0.78	0.41	1.00	0.00	1.00	2935
Black	0.07	0.25	0.00	0.00	1.00	2935
Hispanic	0.06	0.24	0.00	0.00	1.00	2935
Asian	0.06	0.23	0.00	0.00	1.00	2935
Full-time employment	0.49	0.50	0.00	0.00	1.00	2941
Part-time employment	0.08	0.28	0.00	0.00	1.00	2941
Unemployed	0.05	0.23	0.00	0.00	1.00	2941
Bachelor Degree	0.34	0.47	0.00	0.00	1.00	2941
Conservative	0.35	0.48	0.00	0.00	1.00	2804

Notes: This Table summarizes the main covariates of all respondents in experiment 4.

Table A.46: Summary Statistics: Experiment 5 (Many task experiment)

	Mean	SD	Median	Min.	Max.	Obs.
Male	0.46	0.50	0.00	0.00	1.00	5068
Income	52964.44	26194.53	55000.00	5000.00	100000.00	4500
Age	37.25	11.71	34.00	17.00	88.00	5068
Household Size	3.68	1.44	3.00	2.00	13.00	5068
White	0.77	0.42	1.00	0.00	1.00	5068
Black	0.07	0.26	0.00	0.00	1.00	5068
Hispanic	0.05	0.21	0.00	0.00	1.00	5068
Asian	0.08	0.27	0.00	0.00	1.00	5068
Full-time employment	0.51	0.50	1.00	0.00	1.00	5068
Part-time employment	0.11	0.32	0.00	0.00	1.00	5068
Unemployed	0.13	0.34	0.00	0.00	1.00	5068
Bachelor Degree	0.37	0.48	0.00	0.00	1.00	5068
Conservative	0.26	0.44	0.00	0.00	1.00	5042
Number of HITs	8951.11	15446.87	2500.00	750.00	75000.00	5068

Notes: This Table summarizes the main covariates of all respondents in experiment 5.

Table A.47: Summary Statistics: Experiment 6 (Effort Experiment: Weak demand)

	Mean	SD	Median	Min.	Max.	Obs.
Male	0.55	0.50	1.00	0.00	1.00	775
Income	32354.84	16969.25	35000.00	5000.00	85000.00	775
Age	37.50	11.79	35.00	21.00	70.00	775
Household Size	3.71	1.46	3.00	2.00	10.00	775
White	0.74	0.44	1.00	0.00	1.00	775
Black	0.08	0.28	0.00	0.00	1.00	775
Hispanic	0.06	0.24	0.00	0.00	1.00	775
Asian	0.08	0.27	0.00	0.00	1.00	775
Full-time employment	0.54	0.50	1.00	0.00	1.00	775
Part-time employment	0.11	0.31	0.00	0.00	1.00	775
Unemployed	0.13	0.33	0.00	0.00	1.00	775
Bachelor Degree	0.41	0.49	0.00	0.00	1.00	775
Conservative	0.29	0.46	0.00	0.00	1.00	770

Notes: This Table summarizes the main covariates of all respondents in experiment 6.