Inattention, Information Provision, and (Failures of) Revealed Preference

John J. Conlon

August 14, 2025*

Abstract

Economists often estimate preferences by looking how demand for a option changes as its attributes change or as beliefs about them change in response to information. I show in a multi-attribute induced-preference experiment that this method can be substantially biased due to differential inattention. Attention is systematically distorted at baseline, with participants reacting much more to some attributes than others. Further, attention is highly *malleable*: information telling participants about the value of a desirable attribute—even when the information is already known, transparently redundant, and explicitly randomly assigned—greatly increases attention to the attribute it describes and distracts from other attributes. It also boosts average demand for the good, implying that inattention takes the form of neglect rather than shrinkage toward a prior. These results combine to imply that information experiments with passive control groups can be greatly misleading about preferences and can even have the wrong sign: in my experiment, reducing overoptimism about an attribute nonetheless *boosts* demand for its associated good. I show, both theoretically and experimentally, that *active* control-group experiments yield preference estimates with the correct sign but potentially very biased magnitudes. Finally I introduce a novel paradigm—parallel active control-group experiments (PACEs)—that can solve this problem: PACEs yield unbiased estimates of preferences under an equal induced-attention assumption, which is testable (and satisfied) in my experiment.

*Department of Social and Decision Sciences, Carnegie Mellon University: jconlon@andrew.cmu.edu. I thank Chiara Aina, Ned Augenblick, Katherine Coffman, Lucas Coffman, Nicola Gennaioli, Ben Enke, Muriel Niederle, Matthew Rabin, Gautam Rao, Chris Roth, Peter Schwardmann, Josh Schwartzstein, Cassidy Shubatt, Jason Somerville, Andrei Shleifer, and Jeffrey Yang for helpful comments. I am grateful for financial support from the National Science Foundation Graduate Research Fellowship under grant DGE1745303, and the Alfred P. Sloan Foundation Predoctoral Fellowship in Behavioral Macroeconomics, awarded through the NBER. The experiment described in this paper was pre-registered on the AEA registry (ID 0010434). A previous version of this paper circulated under the title "Attention, Information, and Persuasion."

1 Introduction

Revealed preference is a cornerstone of welfare economics: we can learn about preferences—and therefore the welfare impacts of policies—by observing how people choose between alternatives. Increasingly, economists have extended this idea to the realm of beliefs. Naturally occurring or exogenous variation in many important features of choices is often unavailable, but experimenters can nonetheless shift *beliefs* about them by providing appropriate information: e.g., about public policies (Haaland & Roth 2020, Alesina et al. 2023), social stigma (Bursztyn et al. 2020, Roth et al. 2024), or the returns to education (Jensen 2010, Wiswall & Zafar 2015, Conlon 2021). Any resulting changes in choices, it is natural to suppose, reveal agents' preferences for the attributes whose perceived values the information altered.

In this paper, I explore how revealed preference can fail—and how it can be salvaged—when inattention distorts choices and when inattention itself responds to information provision. I first describe a simple model in which agents decide whether to purchase a multi-attribute good but may fail to be fully informed about or fully attentive to some of its features. Comparing how demand in such an environment responds to changes in the good's attributes therefore only reveals some combination of agents' relative preferences for them, their beliefs about them, and their attention to them. These same concerns apply to belief revisions in response to information but with an added complication: information describing an attribute may alter not just what agents believe but also what they attend to. A natural hypothesis is that information directs attention toward the attribute it describes and distracts from others. If so, information will affect choices differently than what we would expect from the resulting belief change and participants' preferences, and it may shift behavior even absent any belief change at all (e.g., if the information tells agents something they already know).

Studying these issues empirically requires an environment where agents' true preferences and beliefs are known. I show that in such a case, choice data reveal what agents attend to: deviations compared to the rational benchmark in how much demand responds to different attributes is a sufficient statistic for agents' attention. This observation motivates my focus on a novel experimental task where such control is possible. Participants repeatedly face a binary choice between two abstract "goods" (see Figure 1 in Section 3 for example decisions). One good ("Option A") offers participants an amount of money with certainty plus a lottery that pays a larger sum with a small probability. The alternative good ("Option B") has six attributes (three types of coins and three boxes whose value depends on their color), each of which contributes to the participant's payment if she chooses that option. Crucially, preferences over attributes are known ex ante, since they all ultimately map onto cash, and participants' beliefs can be measured by asking them about this mapping (i.e., checking their comprehension and memory).[1] Random variation in each good's attributes then allows me to estimate empirical responsiveness to their exogenously changing values, which (combined with knowledge of participants' preferences and beliefs) allows me to identify how attentive they are to each attribute.[2]

I show first that even in this simplified choice environment (compared to real-world economic decisions), the average participant is far from the symmetric-attention "rational" benchmark: equivalent changes in the value of different attributes produce substantially different demand responses. Compared to the attribute participants most respond to, choices react between 25% and 89% less to equal-value changes in other attributes ($p < 0.01$ for all comparisons). These distortions appear despite the vast majority of participants being able, in unincentivized debriefing questions, to correctly describe how the possible values of every attribute would contribute to their earnings. Very similar average distortions appear even among participants who are perfectly informed in this way, confirming that deviations from the rational benchmark stem not from incorrect beliefs (e.g., not knowing how much various attributes are worth) but rather from failures of full attention. Of course, the fact that participants' true preferences are known *ex ante* by design is what allows me to interpret deviations from the rational benchmark as inattention. An analyst lacking this knowledge and instead conducting a naive revealed preference analysis—estimating agents' utility by comparing how choices react to changes in attributes—would draw substantially mistaken conclusions about what participants care about.

How does information provision interact with inattention? After an initial set of baseline choices, a subset of participants begin to receive information telling them about the value of one randomly selected attribute. Recall that most participants are already perfectly well-informed about these values. They are also told explicitly that information is randomly assigned independently of the importance of the attribute to the decision. This information therefore cannot affect choices by changing participants' beliefs, but it nonetheless may alter their choices if it shifts which attributes they focus on. Indeed, I find that this information starkly increases responsiveness to the attribute it describes. These effects are large; responsiveness to an attribute increases by 69% ($p < 0.01$) on average in response to information describing it. I find similar effects (77% increase, $p < 0.01$) when restricting the data to

[1]In principle, Option A's lottery raises the question of risk aversion. I discuss in Section 4 how my results are inconsistent with risk aversion driving any of my conclusions.

[2]Note that I use the term "attention" broadly to mean the weight that attributes receive (also sometimes called "decision weights"). I discuss in Section 2 the connection between this notion and other definitions of attention (e.g., gaze, "top of mind", etc.).

the large majority of participants who already know this information, and thus these effects appear to operate primarily by redirecting participants' attention rather than by changing their beliefs.

These results show that information provision shifts attention, but two additional results shed light on the underlying drivers of attention. First, I find evidence of attention *spillovers*: information about one attribute boosts attention to it in part by decreasing responsiveness to other attributes (by 10% in both the full sample and among participants with correct beliefs, $p < 0.05$ for both comparisons).[3] The total spillover effect, summing across all these other attributes, amounts to 75% of the direct effect. These findings point to an attentional capacity constraint, where information primarily operates by reallocating a (more-or-less) fixed attention budget across attributes. Second, information about an attribute boosts demand for its associated good by about 2 percentage points ($p < 0.01$), despite the fact that information provision is known to be random and thus uncorrelated with the value of the attribute or of the good as a whole. Theoretically, this effect depends on what I call the "default" value: how an attribute is treated when the agent fails to attend to it. The fact that the average effect of information is positive is consistent with this default being *zero*. That is, when failing to attend to an attribute, agents fail to incorporate it into their decision at all, rather than simply relying on its expected or average value. Inattention therefore looks like *neglect*, rather than shrinkage toward a prior.

These results combine to imply that information interventions—especially those with so-called "passive" control groups—can lead to dramatically incorrect conclusions about agents' preferences. The final period of the experiment conducts such an intervention to explore this possibility. Recall that one of the alternatives participants can choose has a lottery associated with it. If they choose this option, they roll five six-sided dice at the end of the experiment, earning an additional bonus if their rolls add up to 12 or less. Participants substantially overestimate the odds of winning this lottery, with the average participant believing she has a 27% chance of winning, whereas the true odds are close to 10%. Nonetheless, I find that information alerting participants to the true odds of winning—and therefore providing bad news relative to their priors—actually *boosts* demand for the option that includes the lottery by 5.1 percentage points ($p < 0.01$). The explanation is that this information, in addition to telling participants about the unfavorable odds, also points their attention toward the lottery. Because it is a positive attribute of the good, which participants otherwise partly neglect, this boosts demand despite the bad news it conveys. Adding evidence to this view,

[3] This distraction result is reminiscent of Altmann et al. (2021), who find that incentivizing one task reduces engagement in another task. My findings show that these sorts of cognitive spillovers arise even within individual decisions and without modifying the objective incentives that participants face.

an almost identical piece of information that describes the lottery but does not provide the odds of winning boosts demand even more (by 9.1 compared to 5.1 percentage points, $p = 0.03$).[4]

Can better-designed information interventions recover preferences? One increasingly common design choice is to use what are called "active control groups." In such experiments, rather than comparing a treatment group that receives information to a passive control group that receives none, instead we compare two treatment groups that only differ in the value the information conveys (e.g., different realizations of a random signal). Typically these active control-group experiments (ACEs) are implemented to avoid differential experimenter demand or inferences about the study across treatments, but it is natural to think that such experiments also hold attention constant (Haaland et al. 2023, Stantcheva 2023). I show that with this assumption ACEs deliver preferences estimates of the correct sign but of the wrong magnitude. The explanation is that preferences are only ever defined in relative terms: how much an agent values a given attribute *compared* to another. While ACEs hold fixed across treatments the attention paid to the information's focal attribute, they do not ensure equal attention between it and other attributes. Unless one knows that attention to some comparison attribute is the same as that toward the information's focal attribute, relative preferences between them are still not identified.

My experimental design allows me to explore the severity of this problem. The information provided about Option B's attributes constitute multiple distinct ACEs: we can estimate how choices among participants receiving information about a given attribute change as the value of that attribute (and therefore the information describing it) changes. I find that the relative preferences one would naively estimate, ignoring attention effects, differ dramatically depending on which ACE we choose to analyze. Suppose we wanted to estimate preferences between Option B's two types of attributes: colored boxes vs coins. Some ACEs in my experiment provide information about a coin, and we can look at how reactive demand is to different values of this coin compared to Option B's boxes. Here we would conclude that participants value boxes at only 31% the level of coins ($p < 0.01$ compared to the true induced preference), because this information shifts attention toward the coin it describes. When we instead implement an ACE informing participants about a box, this result flips: here we would conclude that participants value boxes at 147% the level of coins ($p < 0.01$ compared to the truth, and $p < 0.01$ compared to the coin ACE), because this information shifts focus instead toward boxes. Thus the preference estimates we would derive differ by a factor of almost five depending on which attribute the information describes.

[4]In addition, the difference in effects between the informative and non-informative messages is larger for participants with particularly erroneous beliefs ($p < 0.01$).

These problems with active control-group experiments also suggest a solution: what I call *parallel* active control-group experiments (PACEs). PACEs involve two active control-group experiments: one ACE providing participants information about one attribute, and another ACE providing participants information about a second attribute. We can then compare how reactive demand is to each attribute *when information is describing it.* This comparison identifies agents' true preferences under the assumption that information-induced attention levels are constant across attributes. Some natural possibilities will satisfy this assumption (e.g., if attention is binary and information induces all agents to attend to its focal attribute), and I show that it approximately holds in my experimental context: the preference estimate this method yields in my experiment is within 11% of the true induced preference for blocks vs coins and is statistically indistinguishable from it ($p = 0.27$). This result highlights the usefulness of a controlled environment where I can test this assumption empirically and points toward a potentially useful new experimental method for recovering preferences with information interventions.

This paper contributes to a growing literature on the interaction between attention and beliefs. A relatively small number of papers consider how new information affects attention in addition to beliefs (e.g., Schwartzstein 2014), and those that do often focus on the emotional consequences of redirected attention, such as through anticipatory utility (Golman & Loewenstein 2018, Falk & Zimmermann 2024, Wojtowicz et al. 2025, Bolte & Raymond 2025). A larger literature studies the role attention plays as an input into the belief formation process (e.g. Hanna et al. 2014, Enke & Zimmermann 2019, Enke 2020, Graeber 2023, Gagnon-Bartsch et al. 2023, Bordalo, Conlon, et al. 2025). Relative to this work, I focus on how to recover preferences in light of attention effects. My controlled environment also allows me to cleanly estimate how attention to attributes changes in response to information and what sorts of interventions succeed in recovering preferences.

My results also provide a methodological contribution to the rapidly growing literature in economics employing information interventions. Such work studies preferences for public policies (Haaland & Roth 2020, Alesina et al. 2023), avoiding social stigma (Bursztyn et al. 2020, Roth et al. 2024), climate impact (Imai et al. 2024; Schulze-Tilling 2025), college majors (Wiswall & Zafar 2015, Conlon 2021), features of occupations (Cullen & Perez-Truglia 2022, Conlon & Patel 2025), and much more (see Haaland et al. 2023 for a wider review). My experiment, by studying a controlled lab-experimental environment, shows that we should exercise caution in interpreting responses to information as (only) reflecting agents' preferences. This insight compliments some previous papers studying information provision experiments with apparent "backlash" effects (e.g., Barrera et al. 2020, Alesina et al. 2023, Colonnelli et al. 2024). My results also suggest a new tool (PACEs) for disentangling preferences vs attention in cases where we are unsure of the magnitude (or even the direction) of agents' true preferences.

Next, this paper speaks to applied work studying how people appear to neglect "shrouded attributes," and how reminders about such attributes can affect choices (e.g., Chetty et al. 2009, Brown et al. 2010, Allcott & Taubinsky 2015, Taubinsky & Rees-Jones 2018, Bradley & Feldman 2020). My results show that similar results apply even to aspects of decisions that are not, in any straightforward sense, hidden from decision makers. Further, unlike in my experiment, these studies necessarily estimate relative attention to only a small subset (usually a single pair) of a product's relevant attributes (e.g., price vs taxes), and thus they cannot speak to total attention allocation or distraction. My study also provides a potentially useful method of separating attention vs preferences among attributes whose preferences we may not know *ex ante*.

Finally, there is a large body of work in psychology on priming effects, whereby making certain concepts (e.g., religion, identity, or money-making) more salient can affect people's decisions and attitudes (see Cohn & Maréchal 2016 and Dai et al. 2023 for meta-analyses). My results show that information aimed at changing behavior by correcting beliefs, because it inevitably involves making salient the dimensions the information describes, has effects that echo parts of this literature. Work on priming in psychology, however, tends not to connect such effects to revealed preference analyses or to information provision.

This paper proceeds as follows. Section 2 describes a simple model of inattention-distorted choice to motivate the experimental design. Section 3 explains the experiment in detail, and then Section 4 presents the main results. Section 5 describes and argues against other interpretations of the main results (e.g., experimenter demand, biased priors). Section 6 discusses how my results speak to recent theories of attention in economics, and finally Section 7 concludes.

2 Theoretical Framework

2.1 Setup

I first describe a simple model to motivate the experimental design. Assume an agent decides whether to purchase a good in a given situation s. The good is characterized by a vector of attributes $\vec{a}(s)$ (for example, a car has a price, certain safety features, fuel economy, etc.). Her utility from purchasing the good is linear in these attributes:

$$v(s) = \sum_{k=1}^{K} v_k \cdot a_k(s) \tag{1}$$

Note that while the attributes of the good can depend on the situation s (e.g., different dealerships may offer cars with different features), the agent's preferences v for such attributes are fixed.

I assume that the *perceived* value of the good depends on the agent's beliefs and on which attributes she pays attention to in s. Let $\theta_k(s) \in [0, 1]$ indicate the extent to which the agent attends to dimension k in situation s. To the extent that she does, she incorporates her belief $\hat{a}_k(s)$ about the level of attribute k, which may differ from its true level $a_k(s)$. To the extent that she neglects k, she uses a default value $\bar{a}_k$, which I assume is fixed across situations. Combining these assumptions, I define her attention-weighted perceived valuation of the good, which I denote by u, using equation 2:

$$u(s) = \sum_{k=1}^{K} v_k \Big[\theta_k(s) \cdot \hat{a}_k(s) + \big(1 - \theta_k(s)\big) \cdot \bar{a}_k\Big] \tag{2}$$

Finally, assume that there is some noise $\epsilon \sim F$ such that she purchases the good whenever $u(s) + \epsilon > 0$. Then demand for the good, the probability that an agent purchases it, is $1 - F(u(s))$. For simplicity, I assume that F is uniformly distributed in the relevant range: i.e., $F'(u(s)) = \frac{1}{\kappa}$ for some constant κ.

Note that this basic formulation is in principle compatible with many different theories of what draws attention. For example, attention could be driven (i.e., $\theta(s)$ could be determined) by focusing, relative thinking, salience, memory of the same or similar products/choice scenarios, or noisy information-processing about attributes (Koszegi & Szeidl 2013, Bordalo et al. 2020, 2022; Bordalo, Burro, et al. 2025, Bushong et al. 2021, Gabaix 2019, Yang & Krajbich 2022). In Section 6, I summarize how my results speak to these theories of attention allocation.

My definition of the attention paid to an attribute is also purposely ambivalent about the cognitive process that allows the agent to respond to it. For example, by my definition, an agent could attend to an attribute even without looking at it (e.g., a voice from the next room). They could even attend to an attribute without consciously thinking about it: for example, by my definition, a pedestrian pays attention to the unevenness of a sidewalk because she appropriately (if subconsciously) adjusts her gait to account for it. Although other notions are appropriate for different domains, my definition is appropriate for the research questions I am studying: when do we fail to account for an attribute in such a way that we make the wrong decision? How should policy-makers and economists understand the effects of information interventions in light of potential attention effects?

2.2 Learning about (In)attention

Under what circumstances can we estimate what agents are paying attention to? I begin by considering simple choice data: observing whether (homogeneous) agents purchase a good depending on its attributes. I abstract away from issues of selection bias and consider a case where situations can be randomized (e.g., in an experiment) or where naturally occurring exogenous variation is available.

I begin by introducing some notation. Let $\Delta_{s\to s'}X$ be the change in a variable X if the situation changes from situation s to situation s': $\Delta_{s\to s'}X \equiv X(s') - X(s)$. Then Proposition 1 follows (all proofs in Appendix B):

Proposition 1 *Let s be a baseline situation. For any attribute f, define s_f^+ as a situation where f is one unit higher than in s but where attributes other than f are known to be the same as in s and attention is the same as in s: that is, $\Delta_{s\to s_f^+}a_f = 1$, $\Delta_{s\to s_g^+}a_g = \Delta_{s\to s_g^+}\hat{a}_g = 0$ for all $g \neq f$, and $\Delta_{s\to s_f^+}\theta = 0$. Then,*

$$\underbrace{\frac{\Delta_{s\to s_k^+}D}{\Delta_{s\to s_j^+}D}}_{\textit{Relative Demand Responses}} = \underbrace{\frac{v_k}{v_j}}_{\textit{Preferences}} \times \underbrace{\frac{\Delta_{s\to s_k^+}\hat{a}_k}{\Delta_{s\to s_j^+}\hat{a}_j}}_{\textit{Beliefs}} \times \underbrace{\frac{\theta_k(s)}{\theta_j(s)}}_{\textit{Attention}} \tag{3}$$

As is intuitive, Proposition 1 makes clear that how demand for the good changes as attributes k and j change depends on three objects: how much the agent cares about these attributes (preferences), how much she knows about changes in them (beliefs), and how attentive she is to each of them. Put differently, absent more data, an analyst cannot distinguish whether (say) a lack of demand response as an attribute changes stems from indifference toward this attribute, ignorance of its changing level, or a failure to notice the change.

Proposition 1 also suggests a strategy for studying attention. In an environment where agents' preferences and beliefs are known, relative demand shifts are a sufficient statistic for (relative) attention. This insight motivates the choice in my experiment to study an induced-preference paradigm (where the value of attributes is set by the experiment) where agents are aware of the value of most attributes (and where this knowledge is verifiable).

Information Experiments. Suppose we have a set of situations S, in which the value of the product's attributes may vary. Let I_k be a corresponding information treatment describing the value of attribute k to agents depending on their situation. That is, for each situation $s \in S$, we have a "corresponding situation" $i(s) \in I_k$ where $\Delta_{s \to i(s)} \vec{a} = 0$ but where $\Delta_{s \to i(s)} \hat{a}_k$ and $\Delta_{s \to i(s)} \theta$ may be non-zero: that is, the information intervention does not change the true product attributes, but it may change agents' beliefs about k as well as (potentially) their attention across attributes.

Proposition 1 suggests that we can test whether information experiments shift attention if we have an environment where beliefs and preferences are known *ex ante*, an insight that Proposition 2 formalizes:

Proposition 2 *Suppose attention is constant within S and within I: $\theta^I \equiv \theta(i) = \theta(i')$ for $i, i' \in I_k$ and $\theta^S \equiv \theta(s) = \theta(s')$ for $s, s' \in S$. Let $i(s) \in I_k$ be the corresponding situation of $s \in S$. Assume $s_j^+, s_k^+ \in S$ are defined as in Proposition 1, and assume that beliefs about j are correct in all situations.*

Then the change in responsiveness to k (relative to j) of introducing an information intervention about k is:

$$\underbrace{\frac{\Delta_{i(s) \to i(s_k^+)} D}{\Delta_{s \to s_j^+} D} - \frac{\Delta_{s \to s_k^+} D}{\Delta_{s \to s_j^+} D}}_{\textit{Change in Relative Demand Responses}} = \frac{v_k}{v_j} \left[\frac{\theta_k^S}{\theta_j^S} \underbrace{\left(1 - \Delta_{s \to s_k^+} \hat{a}_k\right)}_{\textit{Belief Effect}} + \underbrace{\left(\frac{\theta_k^I - \theta_k^S}{\theta_j^S} \right)}_{\textit{Attention Effect}} \right] \tag{4}$$

Proposition 2 makes clear that providing information about k can in principle affect how reactive demand is to k through two channels. First, it can have the familiar effect of operating through beliefs: information might boost responsiveness to k because it makes agents aware of the extent to which k differs across situations. But it can also have an attention effect: if agents increase their attention toward k, that will make demand more reactive to that attribute over and above any effect on beliefs. Indeed, if there is no belief effect—e.g., if beliefs about k are already correct absent the information—then we can identify the effect on attention by looking at how demand responsiveness nonetheless changes. My experiment below constructs such a situation for some of the information it provides in order to test for (and quantify) effects on attention.

Notice that the default value $\bar{a}_f$—how agents treat an attribute f to the extent that they are inattentive toward it—does not appear in equation 4. Nonetheless, we will see in the following section that it plays a significant role in understanding responses to certain types of information interventions. How can we learn about this default? Proposition 3 describes how the effect of information on the *level* of demand (rather than the slope of demand as in equation 4) depends on the value of the default:

Proposition 3 *Suppose attention is constant within S and within I: $\theta^I \equiv \theta(i) = \theta(i')$ for $i, i' \in I_k$ and $\theta^S \equiv \theta(s) = \theta(s')$ for $s, s' \in S$. Then,*

$$E\Big[\Delta_{s\to i(s)} D\Big] = \frac{1}{\kappa}\left[v_k \theta_k^I \underbrace{E\Big[\Delta_{s\to i(s)} \hat{a}_k\Big]}_{\textit{Belief Effect}} + \sum_{f=1}^{K} v_f \underbrace{\left(\theta_k^I - \theta_k^S\right)}_{\textit{Attention Effect}} \underbrace{\Big(E\big[\hat{a}_f(s)\big] - \bar{a}_f\Big)}_{\textit{Priors vs Default}}\right] \tag{5}$$

Proposition 3 shows how an information intervention can change average behavior through the familiar beliefs channel but also through an attention channel. First, of course, if the information tends to increase the agent's belief about the level of attribute k—that is, if agents have pessimistic biased priors about k—this will tend to increase demand in the information treatment I_k compared to the control group S (assuming $v_k > 0$). Note that if the agent's beliefs are unbiased, such that information does not change average beliefs about the value of k, then information will have no average beliefs-based effect.

The second term of equation 5 says that, over and above any effect through beliefs, information can boost demand if it tends to increase attention to attributes that are higher than their default value $\bar{a}_f$. Thus any such effects depend crucially on how attributes are treated when they are not attended to. Suppose that beliefs are unbiased, such that average beliefs do not change. By definition, in such a case each attribute will also on average be equal to agent's (correct) priors. If the default is this prior (as in some models of optimized limited attention, see Gabaix 2019), then such information should have no average effect. If instead inattention takes the form of neglect or downweighting (i.e., $\bar{a}_f$ lower than priors or even zero), as in other models (e.g., Koszegi & Szeidl 2013, Bushong et al. 2021, Bordalo et al. 2022), then information can have attentional effects on choices even when it does not focus the agent on especially positive attributes. Simply increasing attention to attributes that are positive *at all* (not necessarily unusually so) can boost demand. The experiment described below constructs an environment where agents are unbiased (and most are perfectly informed) precisely to test between these two possibilities.

2.3 Learning about Preferences

Economists are often interested in estimating agents' preferences. Note that, because choice is invariant to affine transformations of the utility function, the object of interest is always *relative* preferences, or tradeoffs: how much an increase in one attribute is worth in terms of another attribute (i.e., v_k/v_j for two attributes k and j). We already saw from Proposition 1 that simply looking at how demand for a good changes as its attributes change—what we might call the "naive revealed preference" method—in general may fail to recover these preferences, even with exogenous and independent variation in attributes. First, agents may fail to have accurate beliefs about how attributes differ across situations. And second, they may fail to attend equally to all attributes while choosing. Either condition will break the connection between true preferences and choices.

When can information experiments help to recover preferences? Earlier we assumed that we had access to exogenous variation in the levels of all attributes, but this is clearly unrealistic: many attributes (e.g., the college wage premium) cannot be experimentally manipulated. But if agents have inaccurate beliefs, information interventions provide a feasible method of changing the perceived value of these attributes. If information affects beliefs and choices, it is natural to interpret these effects as revealing agents' preferences. With this idea in mind, a large literature has recently developed using information interventions to learn about agents' preferences.

If information provision alters attention in addition to beliefs, (how) can learn about agents' true preferences? I start by defining three types of information interventions. First, consider a "passive control-group experiment" (PCE). A PCE includes both a control group c who receives no information and an information group i_k^+ in which beliefs about attribute k increase by one unit. We will also include a comparison group c_j^+ in which the value of an alternative attribute j is (known to be) one unit higher but where attention is unchanged compared to the control ($\theta(c) = \theta(c_j^+)$). This last group we can interpret as naturally occurring variation in attribute j and will allow us to compare reactions to information about one attribute in terms of reactions to another (i.e., to estimate v_k/v_j). That is, we compare how demand changes in response to the information compared to the benchmark: i.e., we estimate $\Delta_{c \to i_k^+} D / \Delta_{c \to c_j^+} D$.

Second, consider an "active control-group experiment" (ACE). Compared to the PCE, the ACE adds a fourth situation: another information treatment i_k^- in which beliefs about attribute k are now reduced by one unit. Such active control-group designs are often introduced to alleviate concerns about differential experimenter demand between the treatment and control group, but it is natural to also assume that they hold attention constant as well. That is, I assume $\theta(i_k^-) = \theta(i_k^+)$. For comparability, I also assume we have a corresponding baseline condition c_j^- that reduces the (known) value of a_j by one unit without changing attention: $\theta(c_j^-) = \theta(c_j^+)$. We then compare how demand changes in response to this pair of information interventions compared to the benchmark: that is, we estimate $\Delta_{i_k^- \to i_k^+} D / \Delta_{c_j^- \to c_j^+} D$.

Finally, consider "parallel active control-group experiments" (PACEs). PACEs replace the comparison between the benchmark and baseline situations (c_j^- and c_j^+) with a second ACE, now providing information focusing on the alternative attribute j rather than k. That is, we introduce situations i_j^+ and i_j^-, in which beliefs about the value of attribute j are increased or decreased by one unit. Again, I assume that attention is constant across these two situations ($\theta(i_j^+) = \theta(i_j^-)$), though it may differ from attention in the situations providing information about k. We then compare how demand changes in response to each information intervention: that is, we estimate $\Delta_{i_k^- \to i_k^+} D / \Delta_{i_j^- \to i_j^+} D$.

Proposition 4 then describes under which circumstances each method reveals true (relative) preferences:

Proposition 4 *Preference estimates from...*

(i) Passive control-group experiments (PCEs) do not reveal preferences unless attention is symmetric and unresponsive to information. They can have the wrong sign if information affects attention:

$$\frac{\Delta_{c \to i_k^+} D}{\Delta_{c \to c_j^+} D} = \frac{v_k \, \theta_k(i_k^+) + \sum_{f=1}^{K} v_f \; \Delta_{c \to i_k^+} \theta_f \; \left(\hat{a}_f(c) - \bar{a}_f\right)}{v_j \, \theta_j(c)} \tag{6}$$

(ii) Active control-group experiments (ACEs) have the correct sign but wrong magnitude unless attention is symmetric and unresponsive to information:

$$\frac{\Delta_{i_k^- \to i_k^+} D}{\Delta_{c_j^- \to c_j^+} D} = \frac{v_k \, \theta_k(i_k^+)}{v_j \, \theta_j(c)} \tag{7}$$

(iii) Parallel active control-group experiments (PACEs) are correct if attention to the information's focal attribute is constant across experiments:

$$\frac{\Delta_{i_k^- \to i_k^+} D}{\Delta_{i_j^- \to i_j^+} D} = \frac{v_k \, \theta_k(i_k^+)}{v_j \, \theta_j(i_j^+)} \tag{8}$$

As Proposition 4 describes, both passive (PCE) and active (ACE) control-group experiments fail to yield correct estimates of agents' preferences unless attention is equal across attributes and unresponsive to information. With PCEs, such estimates can even be wrong-signed depending on the default values $\bar{a}_f$. We saw a version this result above in Proposition 3. For example, suppose these defaults are zero (inattention taking the form of complete neglect rather than reliance on a prior) and information gives the bad news that a desirable attribute is less positive than expected. If the agent was not previously attending to this attribute, this bad news can nonetheless boost her demand for the good. ACEs eliminate this potential for such wrong-signed responses by holding attention constant within the information treatment groups. But unless the information happens to shift attention to k to the same level of attention that j receives at baseline, estimated preferences will nonetheless be biased.

Parallel active control-group experiments (PACEs) rely on an arguably weaker condition to recover true preferences: that information across experiments induces equal attention to the attribute it describes. In other words, agents must pay the same amount of attention to k while receiving information about k as they do to j while receiving information about j. Some natural possibilities will deliver this result. For example, suppose that attention is binary—$\theta_f \in \{0, 1\}$—but that attributes differ in the probability that agents attend to them. If information guarantees that agents attend to the attribute it describes, then PACEs will recover true relative preferences.

To summarize, whether (and what type of) information experiments can recover preferences depends on three questions: 1) whether there is differential attention at baseline, 2) whether information shifts attention, and 3) whether information-induced attention levels are equal across attributes. As described in Section 2.2, these questions will typically be unanswerable in field settings where either preferences or beliefs are unknown and therefore (in)attention is not identified. But they are testable in a lab-experimental setting where preferences can be induced and beliefs controlled/observed. The next section describes such an experiment.

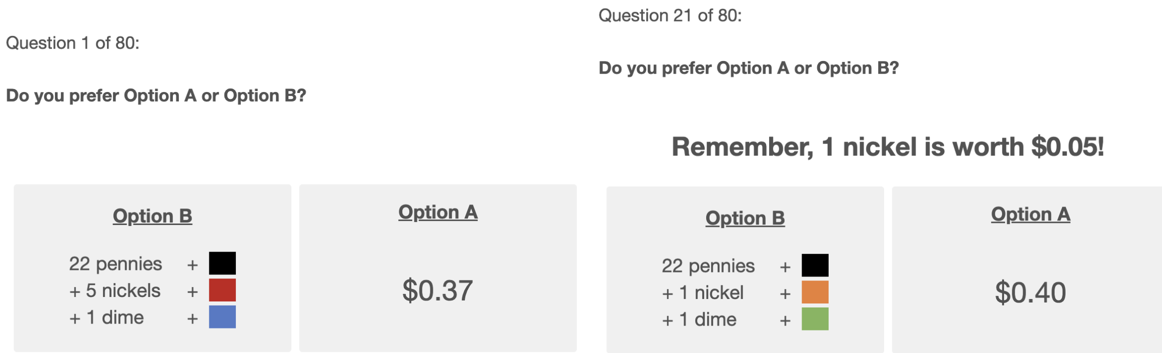
3 Experimental Design

3.1 Option A vs Option B

Participants were recruited through Prolific to participate in an online experiment (see Appendix C for details on recruitment, sample, compensation, comprehension checks, and pre-registration). The main part of the experiment asked participants to repeatedly choose between two options, labeled Option A and Option B, for how their bonus payment would be calculated. They made 80 such choices, and one of these was randomly chosen at the end of the experiment to actually determine their bonus payment. The order in which the two options were presented (Option A on the left and Option B on the right, or vice versa) was randomized across participants but held fixed throughout the experiment. Figure 1 shows screenshots of two such choices.

Option A had two "attributes." First, it listed an amount of money that would, with certainty, be added to participants' bonus if they chose this option. The exact amount (though participants were not told this) was chosen independently across choices from a

Figure 1: Main Experimental Task: Choosing between Options A and B



Notes: This figure gives two examples of the decision screen participants saw when choosing between Options A and B. The left panel is an example of a decision without any additional information being shown. The right panel shows a decision screen for a participant receiving information about one attribute of Option B (in this case, the number of nickels). The left-right placement of Option A vs B was randomized across participants but was constant throughout the experiment.

normal distribution with a mean of $0.40 and a standard deviation of $0.20 (with a minimum of $0.00). Second, if they chose Option A in the decision that was randomly selected to be implemented, they also got to roll five virtual six-sided dice. If the sum of these rolls added up to 12 or less, an extra $1.00 or $2.00 (randomized across participants) was added to their bonus. Right after the instructions page that described this lottery to them, participants were asked their belief about the percent chance of winning such a lottery. The average [median] answer was 27% [20%], significantly higher than the true chance ($p < 0.01$), which is approximately 10%.

Option B had six attributes. I will later look at the impact of providing information about one of these attributes on how responsive participants are to them (more details below). They were therefore designed to be as similar to each other as possible while remaining distinct enough to be considered separately. Three of these attributes were listed numbers of coins (pennies, nickels, and dimes), the value of all of which would be added to their bonus if they chose Option B. There were always 2, 12, or 22 pennies; 1, 3, or 5 nickels; and 0, 1, or 2 dimes. Each of these three values was equally likely to be chosen. Notice that the three coins therefore took on almost the same range of monetary values and could differ from choice to choice by the same amounts (always 0, 10, or 20 cents different). Further, the participant pool was restricted to people living in the US for whom the value of these coins is familiar (which I confirm below).

The other three attributes of Option B were colored boxes (arranged vertically such that there was a top, middle, and bottom box), each of which could take on one of three colors (a different three colors for each box). At the beginning of the experiment, before any other instructions, participants were asked to rank each set of three colors according to "how much you like them." Whichever color they ranked highest then added \$0.20 to their bonus, the one they ranked second added \$0.10, and the one they ranked last added \$0.00. Participants were told this was how these values were assigned. They were also truthfully told that each color was equally likely to appear. Notice that all the colored boxes therefore take on a similar range of monetary values as each other and as the coins, and they could differ from choice to choice by the same amounts (again, always 0, 10, or 20 cents different). The purpose of assigning the values of the colors according to participants' preferences was to make them easy for participants to remember (which I confirm below).

The values of five of the six attributes of Option B were chosen randomly and independently across each choice, with each of the three possible values being equally likely to be chosen. The sixth attribute, randomly selected, was "frozen" at one particular value for the entire experiment. This value was also equally likely to be any of the three possible values for the attribute, but simply did not vary from choice to choice.

3.2 Information about Attributes

Unknown to participants, the 80 choices were divided into four periods, each of which lasted for 20 choices. Periods differed in whether and what type of information participants in different treatment groups were provided while they made their choices. Table 1 summarizes each period for the four different treatment groups.

During Period 1, participants simply chose between Options A and B, as described above, without receiving any additional information. This period was intended to give participants' experience with the decision environment and expose them to the distribution of each attribute's values. During Period 2 (choices 21 to 40), 80% of respondents (Treatments 2, 3, and 4) began to see information at the top of the screen about a randomly selected attribute of Option B (chosen with equal likelihood from among the five non-frozen attributes), which I call the "target attribute." This information told participants how much the target attribute in the current choice was worth. For example, it might read, "Remember, a gold top box adds \$0.20!" or "Remember, 12 pennies add \$0.12!" This message would change as the value of the target attribute changed across choices.

Participants were explicitly told, directly before beginning to choose between Options A and B, that information like this might appear, that any information they were provided was chosen randomly, and therefore that what the information described "is no more likely to be important to your decision" than aspects it did not describe. They then had to

Table 1: Information across Treatment Groups

	Treatment 1	Treatment 2	Treatment 3	Treatment 4
Pre-Decisions	⟵ Identical Instructions ⟶			
Period 1: Decisions 1-20	None	None	None	None
Period 2: Decisions 21-40	None	Target	Target	Target
Period 3: Decisions 41-60	None	None	Target	Alternative
Period 4: Decisions 61-80	⟵ None, Lottery, or Lottery + Odds ⟶			
N	134	239	114	103

Notes: This table describes the distribution of participants across treatment groups and the information presented to each group throughout the experiment. "Target" denotes information about the randomly chosen target attribute of Option B. "Alternative" denotes information about the alternative attribute, which was chosen randomly from the five non-target Option-B attributes. "Lottery" indicates information mentioning the lottery associated with Option A but not its odds. "Lottery + Odds" indicates information that also mentioned the numerical odds of winning Option A's lottery. The first row indicates that instructions were identical to all participants, regardless of treatment group. The row for Period 4 indicates that information about lotteries (or not) was randomly assigned independently of treatment group. The experiment sorted participants into treatment groups at the beginning of the survey, with 20% probability of being assigned to Treatments 1, 3, and 4, and a 40% probability of being assigned to Treatment 2. Variation in sample sizes from this distribution is due to chance.

correctly answer a comprehension question confirming their understanding of this fact before continuing with the survey. This explicit description of the random assignment of information provision shuts down any potential explanation for treatment effects whereby participants infer an implicit recommendation from certain aspects of the experimental manipulation.

During Period 3 (choices 41 to 60), respondents who received no information in Period 2 (Treatmtent 1) continued to see no information. Among participants who received information in Period 2 (Treatments 2-4), Treatment 2 reverted to seeing no information in Period 3. Treatment 3 continued to see information about the target attribute just as they did in Period 2. Treatment 4 was instead shown information about a new attribute (picked at random from the remaining four non-frozen attributes), which I call the "alternative" attribute.

Period 4 did not provide any information about the attributes of Option B. Rather, and within treatment groups, participants were randomized to either receive no information or to receive one of two messages about the lottery associated with Option A. The first message, which I call "Lottery," simply described the lottery (which, though described in the instructions and comprehension checks, was not mentioned on the later decision screens). The second message, "Lottery + Odds," was almost identical but included the numerical odds of winning. In particular, the "Lottery + Odds" message read "Remember, Option A also comes with a 10% chance to win an additional $1 [or $2] prize!" with smaller text at the bottom of the screen telling them the details of how the lottery worked. The "Lottery" message was identical except "10%" did not appear. Because the lottery did not vary from choice to choice, the message remained unchanged at the top of the screen for the entirety of Period 4.

3.3 Beliefs about Attributes

After making all 80 decisions, participants were asked whether they knew how each possible value of Option B's attributes contributed to their bonus payment. In particular, they were first asked how much money a penny, a nickel, and a dime were worth. Reassuringly, 95% of participants get all of these questions correct. Next, for each of the colored boxes associated with Option B (recall there was a top, middle, and bottom box), they were asked to match each possible color to its monetary value ($0.00, $0.10, $0.20). For each box, between 87 and 89% of participants get all three values correct. Seventy-four percent of participants get all 12 questions (three coins, and three values for each of the three boxes) right. Note that this high accuracy appears despite these questions being unincentivized. We see similar levels of accuracy (indeed slightly higher, 75%) among participants who received no relevant information about the attributes during the 80 choices (Treatment 1). This high level of accuracy likely in part reflects the fact that the mapping between colors and money depended on participants' preferences over colors (and thus participants could reconstruct values by thinking about these preferences).

4 Results

4.1 Empirical Strategy

For most of the results below, I estimate variants of equation 9 by OLS:

$$ChoseOptionB_{i,t} = \beta_0 T_{i,t} + \sum_k \beta_k a_{i,k,t} T_{i,t} + \mu_i + \epsilon_{i,t} \tag{9}$$

In the above equation, $ChoseOptionB_{i,t}$ indicates whether participant i chose Option B in decision t. $T_{i,t}$ indicates some treatment status (e.g., whether/what type of information was visible for i during t). $a_{i,k,t}$ is the monetary value of attribute k in that choice, where there are eight possible attributes: the three coins, three colored boxes, the certain payment in Option A, and the value of the lottery in Option A. I define $a_{i,k,t}$ for each attribute such that β_k should be positive for all of them (i.e., I multiply Option A's attributes by negative one). I also scale variables such that coefficients can be interpreted as the effect of increasing the value of an attribute by $0.10. In practice, I often add multiple attributes together (e.g., the value of all the colored boxes, or all non-frozen Option B attributes) to increase power and interpretability. I also typically recenter all attribute values such that they have mean zero. Finally, I also usually include individual-fixed effects μ_i, which could represent person-specific biases toward one or the other option, or heterogeneity in default values $\overline{a_k}$.

Estimating equation 9 lets us answer several questions. First, it tells us how responsive demand for Option B is to the values of various attributes within any given treatment (i.e., any information environment). Recall from Proposition 1 that comparing responsiveness to various attributes reveals a combination of 1) preferences over them, 2) beliefs about changes in them, and 3) relative attention to them. By design, the experiment fixes preferences over attributes, as each attribute of Option B is straightforwardly worth a certain amount of money.[5] In addition, as described in Section 3.3, I measure participants' beliefs about attributes. I can thus ask whether differential responsiveness to attributes is due to differential beliefs or instead to inattention. Similarly, I can ask whether changes in this responsiveness across treatment groups are due either to treatment effects on beliefs or on attention, as in Proposition 2.

Next, the first term in equation 9 tells us how demand responds on average to information about an attribute, pooling across the particular values that attribute takes on. Recall from Proposition 3 that this effect depends both on whether information shifts attention and on what the "default" value is (i.e., how an attribute is treated when it is neglected).

Note that equation 9 has a clear "rational" benchmark: if agents' pay equal attention to all attributes, and if they have correct beliefs about those attributes, then demand responses should all be equal (i.e., $\beta_k = \beta_j$ for all k and j). Further, information should have no effect ($\beta_0 = 0$ and β_k should not depend on treatment).

4.2 Attention at Baseline

Is attention systematically distorted at baseline, i.e., in Period 1 before participants received additional information about any attribute? The left panel of Figure 2 (and column 1 of Table A.I) show estimates of equation 9 (without individual-fixed effects or treatment dummies) for four attributes: Option A's certain monetary value, the subjective value of Option A's lottery, Option B's coins, and Option B's colored boxes. To calculate each participants' subjective value of Option A's lottery, I multiply its monetary prize by participants'

[5] In principle, the presence of the lottery for Option A raises the possibility that risk aversion could affect how participants are willing to trade-off between Options A and B. In practice, as we will see, participants are *more* responsive to the certain payment in Option A than to Option B's attributes, the opposite of what we would expect if risk aversion were a substantial factor in participants' decisions. Since this issue would not substantially affect interpretation any of the main results, I assume participants are risk-neutral.

stated priors about their odds of winning.[6]

The gray dots in the left panel of Figure 2 show coefficients assuming "rational choice": i.e., the expect-payoff maximizing choice given accurate beliefs. Unsurprisingly, we see the expected equal-attention benchmark, with all coefficients close to 0.13, indicating that on average increasing the value of Option B by 10 cents increases the chances that it is the payoff-maximizing choice by 13 percentage points. I cannot reject for the rational benchmark that coefficients across all attributes are equal ($p = 0.79$).

In contrast, we see large differences across attributes when looking at participants actual decisions in this average measure of attentiveness. Compared to Option B's colored boxes, participants are 113% more attentive to the value of Option B's coins and 180% more attentive to Option A's certain value. They appear least attentive to Option A's lottery (which, recall, was not visually represented on each decision screen), as the coefficient on it is only 31% of that on Option B's colored boxes.[7] For each pair of attributes, we can reject that the responsiveness of demand is equal ($p < 0.01$ for all pairwise comparisons).

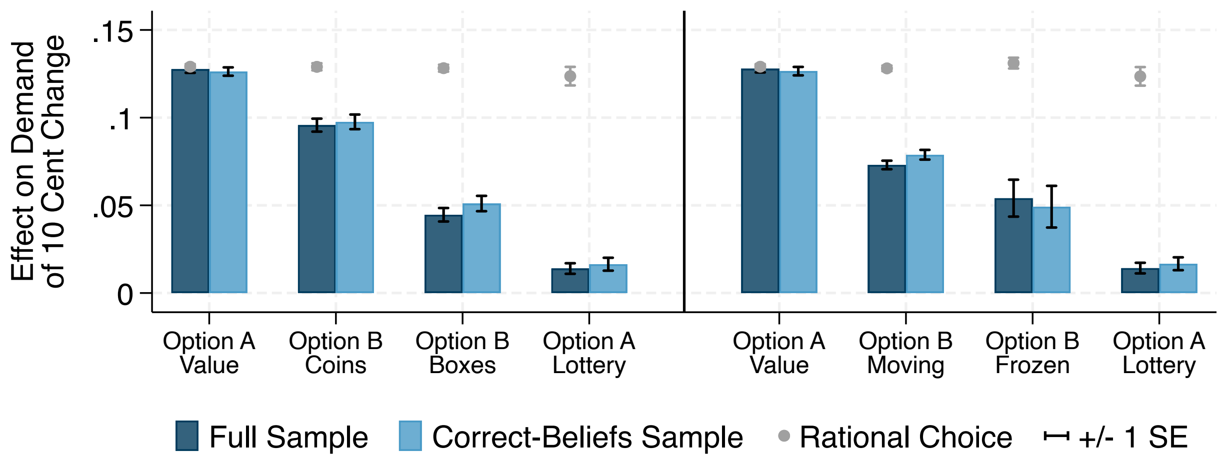
By construction, these differences cannot be due to differences in how participants actually value these attributes. Some differences could, however, be due to misperceptions about how each attribute would contribute to their bonus payment (e.g., not remembering what each colored box is worth). To explore this possibility, the light blue bars in Figure 2 (and column 4 of Table A.I) restricts the analysis to the "correct-beliefs sample," the 74% of participants who correctly reported, in the unincentivized questions at the end of the experiment, the value of each type of coin and the values of each possible colored box.[8] We see similar estimates to those from the full sample: these particpants are 147% more responsive to Option A's certain value, 92% more responsive to Option B's coins, and 69% less responsive to Option A's lottery than they are to Option B's colored boxes. Again, we can reject equality of responsiveness for each pair of attributes ($p < 0.01$). Thus, these differences appear to be driven by differential attention, rather than by mistaken beliefs.

[6] In the main text I use a simple linear probability model, since it makes the fewest assumptions. But Table A.II shows qualitatively identical patterns when I instead estimate a logit regression.

[7] This coefficient could suffer from attenuation bias to the extent that participants' reported beliefs about the lottery are noisy. However, if I instead simply use the objective value of the lottery, whose prize was randomized to be \$1.00 or \$2.00, I find that this has no significant effect on whether participants choose Option B and can reject responsiveness equal to even that of the colored boxes ($p = 0.01$, results available upon request). This estimate does not suffer from attenuation bias (since it does not incorporate participants' potentially noisily measured beliefs). Also notice that because participants on average greatly overestimate the probability of winning the lottery, with full attention they "should" react more to increases in the objective value of the lottery than to increasing the value of the colored boxes by an equivalent amount. Thus, my finding that participants react least to the lottery does not seem to be merely a product of attenuation bias.

[8] Some of these participants, between Period 1 and belief elicitation, saw information telling them about one (Treatments 2 and 3) or two (Treatment 4) attributes of Option B. Table A.III shows similar results even for participants in Treatment 1, who were never provided information about Option B's attributes.

Figure 2: Attention at Baseline



Notes: This figure depicts a subset of the OLS estimates from Table A.I, which estimates equation 9 using the Period 1 decisions of all treatment groups. The dependent variable is whether the participant chose Option B. The independent variables in the left-hand panel are the certain value of Option A, the subjective value of Option A's lottery, the sum of Option B's coins, and the sum of Option B's colored boxes. The subjective value of the lottery is calculated by multiplying the prize for winning the lottery with each participants' prior belief about their odds of winning (winsorized at the 90th percentile). The independent variables in the right-hand panel are identical except the Option-B attributes instead include the sum of the five changing attributes and the attribute that was frozen throughout the experiment at its initial value. Dark blue bars show estimates including all participants, while the light blue bars show estimates including only participants who correctly respond to unincentivized questions at the end of the experiment about the monetary value of each coin and every possible colored box. Whiskers show robust standard errors, clustered at the individual level. Table A.I shows the full regression results for these specifications.

A natural question is what drives these attentional differences at baseline. One possibility is that some differences arise due to differences in attributes' associated payoffs within a given decision. Option A's certain value tended to be larger than that of other attributes; if people pay more attention to attributes during decisions in which they take on more extreme values, that might explain why on average Option A's certain value draws more attention. Columns 2 and 5 of Table A.I explore this possibility by adding quadratic terms to the regression specification.[9] Though the estimates are not always statistically significant, the coefficients on the squared terms show if anything the opposite pattern. When Option B's attributes are more valuable, their marginal contribution to choosing Option B is smaller (negative coefficients). Conversely, when Option A's attributes are more valuable, their marginal contribution to choosing option A is smaller (positive coefficient).[10] This result suggests that if anything there is diminishing sensitivity to the value of attributes, so range effects cannot explain the average differences between them.[11] Further, note that Option B's coins and colored boxes have almost identical ranges of values, so the difference between them cannot not be due to such effects.

Next, recall that a randomly chosen attribute of Option B is frozen at its initial value throughout the whole experiment. The right panel of Figure 2 (and Column 2 of Table A.I) show that participants are 26% less responsive to this attribute than to the attributes that are changing for them from decision to decision ($p = 0.08$). We see similar relative inattention among the correct-beliefs sample (38%, $p = 0.02$). I return to these facts in Section 6 when I discuss implications of my findings for models of attention.

4.3 Attention in Response to Information

I now turn to the question of whether information shifts attention in addition to (or even without) changing beliefs. Given that the large majority of respondents are able to recall how each attribute's possible values contribute to the total value of Option B, and

[9]For Option B's coins and colored boxes, I square each component attribute (e.g., the value of the dimes or the middle box) and then add up these squared values. This allows me to test whether extreme individual attributes draw more attention. This is equivalent to regressing choices on the value of each individual attribute and its squared value but with the restriction that there be equal coefficients across coins and across colored boxes.

[10]The fact that positive coefficients on Option-A attributes implies less sensitivity to more extreme values can be understood as follows. As the square of Option A's certain value gets larger, this boosts demand for Option B. Thus, though the main effect of making Option A more valuable of course reduces demand for Option B, this marginal effect favors B more as the level gets larger. Here, however, we should be somewhat cautious, as the rational benchmark is not that these coefficients be zero (column 8 of Table A.I).

[11]In addition, unlike in the analyses to follow, Table A.I does not recenter each attribute to have mean zero. Thus, the coefficients on the main effects can be interpreted as the marginal effect of increasing each attribute from a value of zero. These main effects look qualitatively similar those in column 1 where quadratic terms were excluded.

that they know information is randomly assigned, it might be natural to think that it should have little effect. However, recall from Propositions 2 and 3 the possibility that information might affect choices even without any effect on beliefs if it changes the relative attention that agents pay to different attributes.

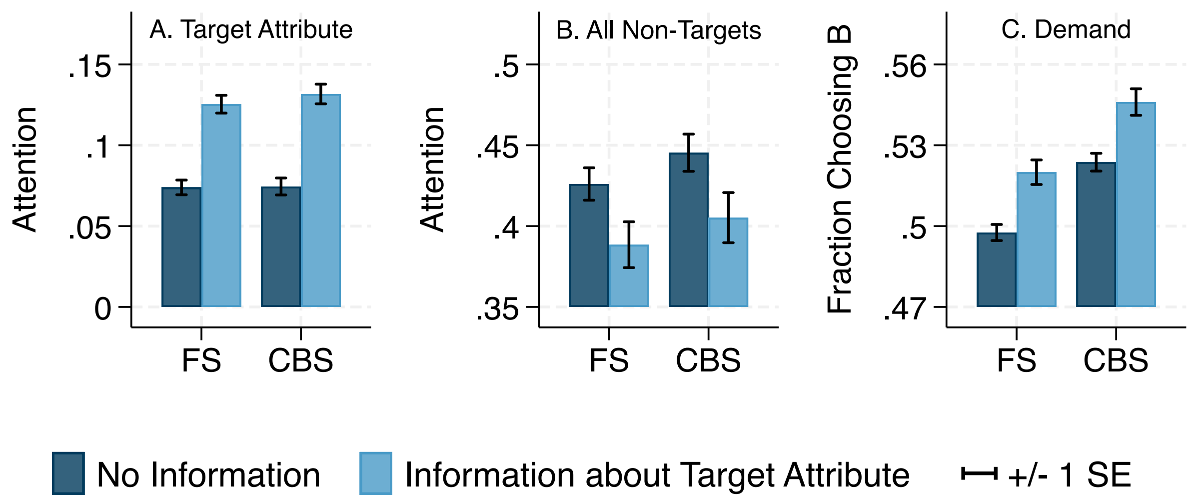
To investigate this possibility, we can compare participants in treatment groups 2-4, who received information in Period 2 about a randomly selected attribute of Option B, to those in Treatment 1, who continued to see no additional information. In practice, the experiment implemented these treatments by choosing, for each person regardless of treatment group, a random attribute of Option B to be the "target" attribute. In Treatments 2-4, participants then began to receive information about this attribute in Period 2. I can therefore compare responsiveness to this target attribute depending on whether participants were (Treatments 2-4) or were not (Treatment 1) receiving information about it.

Figure 3 summarizes OLS estimates of equation 9 (the full estimates are reported in Table A.IV), pooling data from all participants for Periods 1 and 2. In these regressions, the included attributes are the certain value of Option A, the target attribute of Option B, and the non-target attributes of Option B (all summed together). I interact these attributes with a dummy for receiving information about the target attribute (i.e., being in Treatments 2-4 during period 2). We see in the leftmost pair of bars in Panel A (and in column 1 of Table A.IV) that the information has a large effect on responsiveness to the target attribute, increasing by 69% the attention participants pay to it (from 0.074 to 0.125, $p < 0.01$). We see similarly sized effects if we restrict the sample to the "Correct-Beliefs Sample" (right pair of bars in Panel A, and column 4 of Table A.IV), the respondents who correctly identify how each attribute of Option B contributes to their bonus in the unincentivized questions at the end of the experiment. This result is consistent with the information primarily operating by changing how much attention participants pay to different attributes, rather than through its effect on their beliefs.

Columns 2-3 and 5-6 of Table A.IV split the sample by whether the target attribute was a coin or colored box. We see that effects are larger for information about the colored boxes ($p < 0.05$ for both the full and correct-beliefs samples), suggesting that attention effects are larger when baseline attention is lower. However, even information about coins significantly affects the difference in responsiveness to the target attribute ($p < 0.05$ for both samples). Because it is obvious (and the later unincentivized questions confirm) that participants are already perfectly aware that, say, two dimes are worth \$0.20, the most natural interpretation of these results is that this information changes participants' choice by shifting what they attend to.

These results show that information boosts the attention participants pay to the attribute

Figure 3: Effects of Information about Target Attribute



Notes: This figure summarizes OLS estimates of equation 9 using the Periods 1 and 2 decisions of all treatment groups. The dependent variable is whether the participant chose Option B. The independent variables are the attributes of Options A and B interacted with a dummy variable for receiving information about the target attribute (i.e., being in Treatments 2-4 during Period 2). Panel A shows the coefficient on the target attribute. Panel B shows the sum of attention to all non-target attributes. Panel C shows the fraction choosing Option B depending on treatment, controlling for the other variables in the regressions. The left-hand and right-hand pairs of bars within each panel shows estimates from the full sample (FS) and correct-beliefs samples (CBS), respectively. Whiskers show robust standard errors, clustered at the individual level. See Table A.IV for the full regression results and more details on the specification.

it describes. Does it also have effects on other attributes? Panel B of Figure 3 shows the sum of the coefficients on all other attributes for participants who are and are not receiving information about the target attribute. We see that this measure of the total attention paid toward the other attributes in the decision significantly declines when information directs attention toward the target attribute ($p < 0.05$ for both the full and correct-beliefs samples). In percent terms, this decline is smaller (about 10%) than the direct effect on the target attribute, but it applies to many more attributes. Thus in absolute terms the total spillover effect is comparable to the effect on the target attribute (-0.038 compared to 0.051). The effect on "total" attention, summing the effects on the target and non-target attributes, is therefore statistically indistinguishable from zero ($p = 0.43$).

The effects described so far concern how information affects the responsiveness (i.e., *slope*) of demand to various attributes. Panel C of Figure 3 shows that information about the target attribute also significantly boosted the average *level* of demand for Option B by 2.2 p.p. ($p < 0.01$ for both the full and correct-beliefs samples).[12] These effects appear despite the information being uncorrelated with the value of the attribute it described. That is, on average the treatment provided neutral information (which, in any case, most participants already knew) about the value of the target attribute. This result is consistent with the default value—how an attribute is treated when it is not attended to—being *zero* (or at least, less than participants' priors) rather than being agents' priors about its expected value.[13] Consistent with this effect operating through attention, these effects are larger for information about colored boxes than about coins ($p = 0.05$ and $p = 0.10$ for the full and correct-beliefs samples), corresponding to the larger boost in attention for these attributes compared to coins.

4.4 Implications for Recovering Preferences

4.4.1 Naive Revealed Preference

Proposition 1 showed that comparing demand responses to different attributes can fail to recover true relative preferences—even under perfect information—if agents fail to pay equal attention to them. This is precisely what we saw in Figure 2: there are large and systematic differences across attributes in how sensitive demand is to equal-value changes in their levels. An analyst neglecting this fact—employing what we might call a "naive revealed preference method"—would arrive at substantially biased estimates of participants'

[12]The values of each attribute in Table A.IV are recentered to have mean zero, so the main effect of information can be interpreted as the effect at the mean of these values.

[13]Note that, by this point in the experiment, all participants had made at least 20 previous choices, and so had experience with the distribution of values that each attribute took on.

preferences. Note that these failures of full attention appear in my experiment despite none of the attributes being "hidden" or "shrouded". Instead, some attributes simply jump out at (i.e., are salient to) participants or are easier to process. Further, these differences in baseline attention appear systematic and (approximately) in line with some existing theories of attention allocation in economics (see Section 6 for a more detailed discussion).

4.4.2 Passive Control-Group Experiments

We have seen that 1) attention at baseline is systematically distorted, 2) that information shifts attention toward the attribute it describes and away from other attributes, and 3) that the default value appears to be less than participants' priors. Recall from Proposition 4(i) that these are precisely the conditions under which passive control-group experiments (PCEs) can yield not only incorrect estimates of participants' preferences, but estimates that have the wrong sign. Period 4 of the experiment was designed to emphasize and explore this possibility. Recall from Table 1 that during Period 4, participants were randomly (and independently of their earlier treatment group) sorted into three groups. A third of participants saw no additional information in Period 4. Another third were shown the "Lottery" message, which informed them that Option A also came with a lottery that added \$1 or \$2 to their bonus if their roll of five 6-sided dice add up to 12 or less. The final third of participants received the almost identical "Lottery + Odds" message, which additionally included the fact that such a lottery pays off about 10% of the time, much lower than participants' priors (mean 27%, median 20%, both significantly different from the truth at $p < 0.01$).

What effect should we expect the "Lottery + Odds" message to have on demand for Option A? As shown in equation 5, there may be two competing effects. First, in one sense this information clearly conveys bad news about Option A: it should reduce participants' beliefs about the value of the lottery, and hence of Option A. But second, if participants would otherwise fail to pay attention to the lottery, the information's attentional effect depends on the lottery's default value $\overline{a}$. The results in 4.3 suggested that this default is zero. If so, then boosting attention toward the lottery could nonetheless boost demand for Option A by pointing attention toward one of its positive attributes (even if it is not as positive as participants' priors suggested).

Table 2 shows OLS estimates where the dependent variable is whether participants chose Option A (which included the lottery), pooling all decisions across all periods of the experiment. I regress this variable on the certain value of Option A, the total value of all six of Option B's attributes, individual-fixed effects, an indicator variable for whether participants saw the information about the lottery that did not include odds, and a similar indicator for the information that also included the odds of winning. In column 1, we see that the "Lottery + Odds" message significantly *boosted* demand for Option A by 5.1 p.p., ($p < 0.01$), despite delivering bad news about it for the average participant.

We can further disentangle the beliefs and attention effects of the "Info + Odds" treatment by comparing it to the "Info" treatment, which did not change beliefs about the lottery's odds. Column 1 of Table 2 shows that this message boosted demand for Option A by even more (9.1 p.p., $p < 0.01$). This result suggests that the "pure" attention effect is quite large, enough to countervail the significant beliefs effect of 4.0 p.p. (9.1 p.p. minus 5.1 p.p., $p = 0.03$).

In column 2, I additionally interact these indicators with the error (beliefs minus truth) in participants' priors about the odds of winning the lottery. For participants who do not receive information about the odds of winning, we see directionally larger effects for respondents who overestimated the odds of winning by more ($p = 0.20$). In contrast, for participants who also learned the true odds of winning, the interaction term is negative and statistically significant ($p < 0.01$), as we would expect from the information correcting misperceptions. The main effects of both interventions—the effect for participants whose beliefs are correct, which we can interpret as the pure attentional effect—are large and positive (around 8 percentage points, $p < 0.01$, for both), and statistically indistinguishable from each other ($p = 0.72$). Note also that the fact that the "Lottery + Odds" message boosted demand for Option A even for participants whose beliefs were already correct (the interacted main effect in Table 2) is particularly stark evidence that participants' default is less than their priors, consistent again with inattention taking the form of neglect. Similarly, the fact that the simple "Lottery" message boosted demand for Option A without altering beliefs about the odds of winning is again inconsistent with participants' default being their priors.

Note that, in my controlled experimental environment, it is obvious that participants value greater odds of winning the lottery (i.e., that their true preference is positive). This negative effect of information must therefore be operating through changing attention (and the above results corroborate that interpretation). But in field settings and less controlled environments, where we may not know the direction of agents' preferences, it will be difficult to know whether the direction of responses to information are at all reflective of agents' preferences. These results highlight how PCEs can badly fail to uncover agents' preferences in the presence of malleable inattention.

4.4.3 Active Control-Group Experiments

Recall from Proposition 4 that active control-group experiments (ACEs), by keeping attention across attributes fixed between treatment groups, avoids wrong-signed estimates

Table 2: Effect of Information that Changes Both Beliefs and Attention

	(1)	(2)
Lottery Info without Odds	0.091***	0.077***
	(0.013)	(0.017)
Lottery Info including Odds	0.051***	0.085***
	(0.013)	(0.017)
Lottery Info without Odds X Error in Prior		0.099
		(0.076)
Lottery Info including Odds X Error in Prior		-0.217***
		(0.074)
Option A Value	0.128***	0.128***
	(0.002)	(0.002)
Option B Total Value	-0.076***	-0.076***
	(0.002)	(0.002)
Observations	47,200	47,200
Individuals	590	590
R^2	0.50	0.50
p-value: Main Effects of Information Equal	0.03	0.72
p-value: Interactions Equal		0.00

Notes: This table shows OLS regression estimates, pooling data from all Treatments and all Periods of the experiment. The dependent variable is a dummy indicating whether the participant chose Option A (which included the lottery). I regress this variable on the certain value of Option A ("Option A Value"), the sum of all six of Option B's attributes ("Option B Total Value"), individual-fixed effects, and dummy variables indicating whether the participant was shown information mentioning the lottery, where this information either came without numerical information about the odds of winning ("Lottery Info without Odds") or with such information ("Lottery Info with Odds"). In Column 2, I additionally interact these dummy variables with the error (belief minus truth) in participants' previously reported beliefs about the odds of winning the lottery (winsorized at the 90th percentile). Robust standard errors, clustered at the individual level, are presented in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

of relative preferences. But in the presence of differential attention, a single ACE will in general still generate biased estimates of relative preferences. My experiment allows me to investigate this empirically. Notice that Period 2 of the experiment, in which participants were shown information about an attribute across multiple scenarios, constitutes an ACE. That is, we have random variation among participants receiving information about value of the attribute it describes.

We can sort participants into three groups: those who received no information in Period 2 (treatment 1), those among treatments 2-4 who received information about a coin during Period 2, and those who during that period received information about a colored box. Panel A of Figure 4 shows OLS estimates of the equation 10 for each of these groups' Period-2 decisions:

$$\begin{aligned} ChoseOptionB_{i,t} = \alpha_0 + \alpha_1 OptionA_{i,t} + \alpha_2 NonInfoBoxes_{i,t} + \alpha_3 NonInfoCoins_{i,t} \\ + \alpha_4 InfoAttribute_{i,t} + \mu_i + \epsilon_{i,t} \end{aligned} \tag{10}$$

In equation 10, $OptionA_{i,t}$ is the certain value of Option A for participant i in choice t, $NonInfoBoxes_{i,t}$ and $NonInfoCoins_{i,t}$ are the total value of all colored boxes and coins that the participant is not receiving information about, $InfoAttribute_{i,t}$ is the value of the attribute (if any) that the participant is receiving information about, and μ_i are individual-fixed effects.

Figure 4 summarizes OLS estimates of equation 10 (columns 1-3 of Table A.V show the full estimates). The left two bars of Panel A show that participants in treatment 1, who receive no information during period 2, are 53% as reactive to equal-value changes in the value of boxes compared to coins. An analyst employing what I have called the naive revealed preference method—comparing such responsiveness absent information as reflecting only participants true preferences—would therefore wrongly conclude that agents' preferences for boxes are that much much lower than for coins. Panel B plots this "estimate" for the naive revealed preference method, which is significantly different from the true induced (equal) preferences ($p < 0.01$). This finding echoes the baseline-attention result from Figure 2.

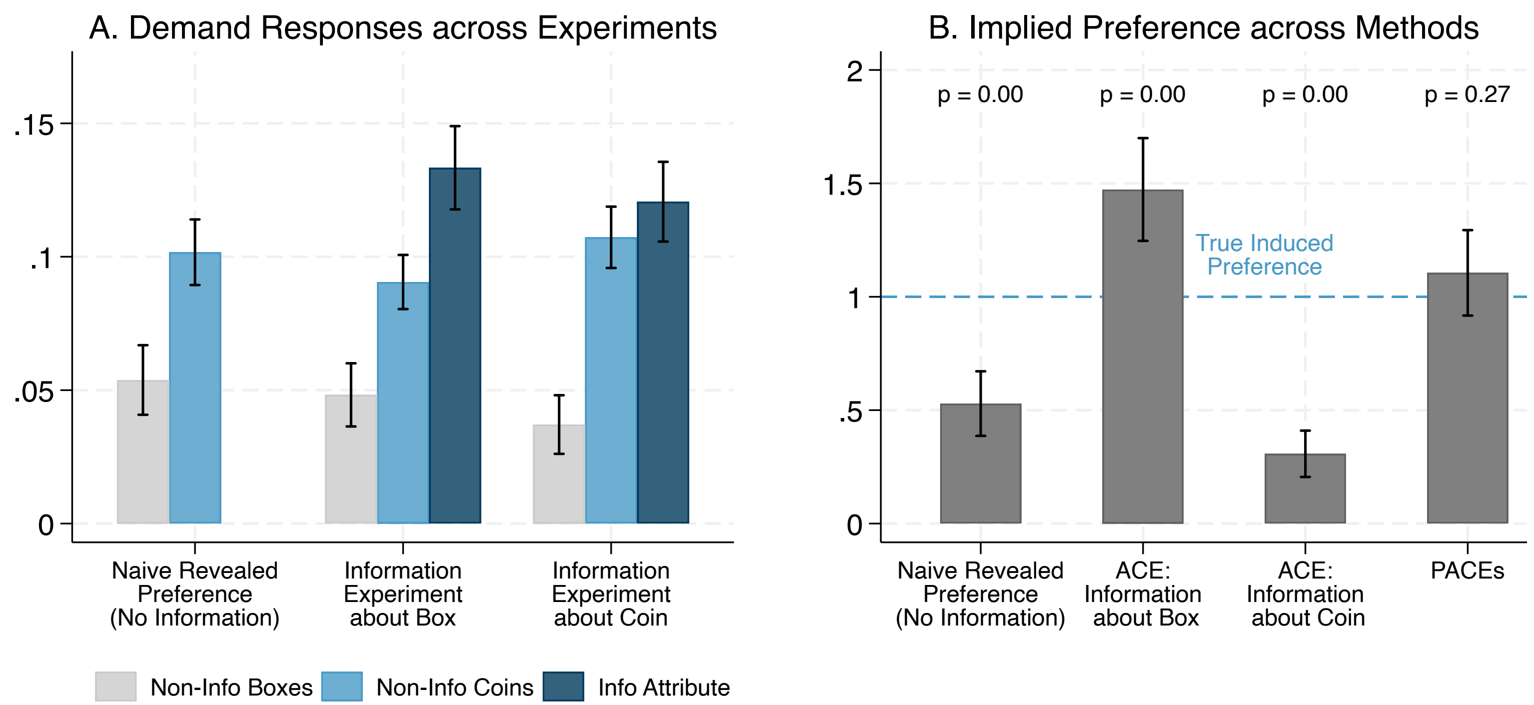
What would an analyst using a single ACE conclude about participants' preferences? The answer depends heavily on which attribute she chooses to (or must) provide information about. The two sets of bars on the right in Panel A of Figure 4 show, as we have seen, that participants become more reactive to the attribute they are receiving information about. The second bar in Panel B assumes the analyst uses an ACE providing information about a colored box to learn about preferences for it compared to coins. That is, it shows the coefficient on the target box divided by the coefficient for coins for such participants. We see that, because responsiveness to the target attribute is much larger with information than without, the "estimated preference" in this case now erroneously ascribes to participants a 47% greater preference for boxes than for coins ($p < 0.01$).

Had the analyst instead chosen to provide information about a coin (final set of bars in Panel A of Figure 4), she would have drawn the exact opposite conclusion. Recall that without information participants already pay more attention to coins than boxes. The fact that information boosts attention to its described attribute and reduces attention to other attributes exacerbates this baseline difference. Estimated preferences for boxes relative to coins shift from 53% under the naive revealed preference (NRP) method to 31% when using an ACE informing participants about a coin ($p < 0.01$ compared to NRP, and $p < 0.01$ compared to the true induced preferences). Estimated preferences under ACEs are therefore extremely sensitive to the choice of which attribute to provide information about: estimated preferences are almost five times as large (479%, $p < 0.01$) if information describes a box than if it describes a coin.

Do parallel active control-group experiments (PACEs) solve this problem and recover true preferences? Recall from Proposition 4 that they do so if attention to information's focal attribute is constant across experiments. We can test for this condition in my experiment by comparing responsiveness to the target attribute across the two types of ACEs (for coins and for boxes). We see that the dark blue bars in Panel A of Figure 4 are approximately equal to each other (0.133 vs 0.121, $p = 0.25$). Taking their ratio (final bar of Panel B), we see that this method therefore approximately recovers participants true preferences, estimating that participants value boxes an insignificant 10.5% ($p = 0.27$) more than coins. PACEs therefore succeed in this context in revealing participants preferences despite (and, indeed, because of) the fact that information shifts underlying attention to attributes.

Figure 4 shows results from all participants. Figure A.I (and columns 4-6 of Table A.V) show analogous results restricting participants to the correct-beliefs sample, where we see very similar results. One difference in the PACEs estimate is now marginally significantly different from the true induced preference ($p = 0.07$), but it is still much closer to it than (and statistically distinguishable from) the the other three methods ($p < 0.05$ for all comparisons).

Figure 4: Recovering Preferences



Notes: Panel A of this figure summarizes OLS estimates of equation 10 (see columns 1-3 of Table A.V for the full estimates). The first pair of bars shows estimates of α_1 and α_2—responsiveness of demand for Option B to the value of its coins and colored boxes—among participants in Period 2 who received no information (treatment 1). The next three bars in Panel A show estimates of α_1, α_2, and α_3 for participants (in treatments 2-4) who received information about a coin in Period 2. The final three bars in Panel A show analogous estimates for participants who received information abotu a colored box in Period 2. Panel B then computes preference estimates assuming different methods. The first bar shows estimated preferences for blocks vs colored boxes assuming "naive revealed preference": i.e., dividing the estimate for α_1 by α_2 among those who received no information in Period 2 (gray and light blue bars in leftmost pair of bars in Panel A). The next bar shows the quotient of α_1 and α_3 among those receiving information about a coin (dark blue divided by light blue from middle three bars in Panel A). The next bar shows the quotient of α_3 and α_2 among those receiving information about a block (gray divided by dark blue from right three bars in Panel A). The final bar in Panel B shows estimates from the parallel active control group experiments (PACEs) method: dividing α_3 among those receiving information about a box (dark blue from middle bars in Panel A) by α_3 among those receiving information about a coin (dark blue from rightmost bars in Panel A). Whiskers show 95% confidence intervals from robust standard errors clustered at the individual level.

4.5 Additional Results

4.5.1 Dynamics of Attentional Effects

Period 3 of the experiment allows me to investigate the stability and longevity of the attention effects documented above. Figure A.II and Table A.VI show estimates of equation 9 using data from only Period 3, where the four attributes are the certain value of Option A, the target attribute of Option B, the "alternative" attribute of Option B (which Treatment 4 sees information about in Period 3), and the remaining attributes of Option B (all summed together). I estimate this regression separately for each treatment group. We see that having received information about the target attribute still has a significant effect on attention paid to it, even when this information is no longer visible (Treatments 2 vs 1, 0.119 vs 0.064, $p < 0.01$). This increased attention is almost identical to the attention paid to the target attribute when the information is still visible (Treatments 2 vs 3, 0.119 vs 0.122, $p = 0.81$). In contrast, when information about a new attribute begins appearing (Treatment 4), the effect of the previous information disappears entirely: attention to the target attribute reverts to a very similar level as if participants had never received the information (0.064 vs 0.067 in Treatments 1 vs 4, $p = 0.55$) and much less than if the new information had not appeared (0.119 vs 0.067 in Treatments 2 vs 4, $p < 0.01$). We also see a large and significant effect on attention paid to the alternative attribute in Treatment 4 (0.120 vs 0.077 in Treatments 4 vs 1, $p < 0.01$), as expected.

Taken together, these results suggest that, while information can have large effects on attention and that these effects can outlast the information itself, they are also quite fragile. I interpret this result as another manifestation of distraction: just as information boosts attention to the attribute it describes by decreasing focus on other attributes, so the attentional impact of new information comes at the expense of the effect of previous information. These results also suggest that taking advantage of the attention-boosting effects of information provision (e.g., to estimate preferences using PACEs) requires measuring choices very close to the time of (and ideally simultaneously with) information provision.

4.5.2 Analyzing Open-Ended Text Responses

At the very end of the experiment, participants were asked to describe in their own words how they tended to choose between Options A and B. I then have a large language model categorize their responses to shed some (suggestive) light on participants' choice processes (see Appendix C for details on the question and categorization procedure). The average (median) respondent wrote 23 (18) words in response to this prompt, so these answers provide short summaries of what strategies participants felt they were employing. However, as with many (unincentivized) open-ended questions, the level of engagement from respondents shows considerable variation: 22% of participants write less than ten words.

I find, reassuringly, that the large majority of participants (66.9%) mention attempting to add up the monetary value of each option before choosing. At the same time, many participants are aware that they are doing so only imperfectly, with 42.5% mentioning that they were merely trying to identify or guessing (rather than knowing) what the right option to choose was. Perhaps surprisingly given the large effect that they had on participants' choices, only 1.2% (7 participants) mention the information treatments at all in these open-ended questions. Finally, no participant makes any mention of reacting to what they perceived the experimenter to want them to do, or of interpreting the information provided as an implicit suggestion of the right course of action, facts I return to in Section 5.

5 Ruling out Potential Confounds

To summarize the results in Section 4, I interpreted four results in terms of attention. First, at baseline (i.e., absent any additional information), participants' responsiveness across attributes deviates from a rational full-attention benchmark. Participants are less attentive to the colors and coins of Option B than to the certain value of Option A, and they are least attentive to Option A's lottery. Second, providing information (even information that participants already know) boosts attention paid to the attribute it describes. Third, this boosted attention to the described attribute comes at the cost of distracting attention away from other attributes. Finally, randomly assigned information about one of Option B's attributes boosts demand for that option, implying that the default value for attributes is below participants' priors about them, consistent with inattention as neglect. This latter finding suggests that even bad news about an attribute (relative to participants' priors) can nonetheless boost demand for its associated option, an implication we saw borne out in Section 4.4.2.

In this section, I consider potential confounds to my interpretation of these results and describe how the design and findings either rule them out or render them unlikely.

Pessimistic Priors

One might worry that incorrect priors about the distribution of attribute values might drive some of these results. For example, suppose a participant believes that one of Option B's attributes has a lower average value than it in fact does. Suppose further that she only incorporates the value of that attribute into her choice when information about it is being provided, and that otherwise she relies on her biased prior. If so, she will tend to react to accurate information about it by increasing her demand for Option B. She might do so even if the default value she uses absent attention is equal to her prior. Thus, a positive average effect of information on demand for Option B would not be a result of inattention taking the form of neglect.

Several aspects of the experimental design and results rule this explanation out. First, for some attributes, participants are explicitly told the correct priors about the distribution of values. In particular, the instructions told participants that each color (for each of the colored boxes) was equally likely to be chosen. Despite this, information about these attributes boosted demand for Option B more than information about other attributes (Table A.IV, columns 2-3 and 5-6), the opposite of what we would expect if pessimistic priors were driving these responses. Second, participants had extensive experience in the first period of the experiment (20 decisions with randomly generated attribute values) during which they encountered the true distribution and could update any incorrect priors. It is only after this initial period that the experiment begins to provide information, and thus pessimistic priors would have to persist despite repeated feedback.

Finally, the results in Section 4.4.2 provide particularly clear evidence against the view that participants' default value is their prior. In particular, we saw that information describing Option A's lottery boosted demand for Option A *even among those who were not told the true odds of winning* (Table 2, column 1). Such participants must have relied on their prior about the lottery's odds, and so if the default were their prior this treatment should have had no effect. In addition, we also saw positive effects on demand for Option A of information describing the odds of winning *even among participants who already had correct priors* (Table 2, column 2). If participants were already relying on their prior, information confirming this belief would have had no effect.

Information as an Implicit Suggestion

Next, a potential worry is that participants interpreted the information, despite it containing no wording suggesting they ought to, as an implicit recommendation of the action they should take or of the attributes that happen to matter more in the decision. This possibility is ruled out by the design of the experiment, as participants are explicitly told that any information is about randomly selected attributes and therefore not a signal of what they should react to. A comprehension question immediately prior to the beginning of the choice task required participants to confirm their understanding of this fact, including the implication that information should not alter their beliefs about which attributes are most important. The fact that almost no participants (1.2%) mentioned the information, when asked at the end of the experiment how they decided between Options and B (see Section

4.5.2), is consistent with this understanding that the information did not serve as an implicit endorsement of any action.

Experimenter Demand

Next, one might worry that the results could be driven by experimenter demand: i.e., that participants might attempt to anticipate the researcher's hypotheses and then modify their behavior to confirm these hypotheses out of sense of altruism. Several observations speak against this concern. First, there are a fairly large number of results consistent with some (but not all, see Section 6) behavioral-economic theories that participants would need to anticipate and then conform their behavior to. Some of these are subtle enough that participants' anticipating them seems quite implausible. For example, participants' boost their responsiveness to attributes the information describes *and* reduce their responsiveness to attributes it does *not* describe (Figure 3); participants' attention to described attributes reverts to its no-information baseline level *only* when it is replaced by information about a new attribute (Figure A.II); and participants' increase their demand for options whose attributes the information describes *unless* their priors about that attribute are sufficiently pessimistic (Table 2).

Second, and more fundamentally, there is no empirical evidence that experimenter demand is a salient concern in economics lab experiments. De Quidt et al. (2018) find that explicitly asking participants to change their behavior as a favor to the experimenters has effects around 0.13 standard deviations across 13 different standard experimental tasks. Similarly, Winichakul et al. (2024) study four decision domains among three different subject pools and find no evidence that even the strongest experimenter-demand treatments change behavior enough to spuriously produce large effects. Crucially, these studies estimate *upper bounds* on experimenter demand effects, where the change in behavior required to benefit the experimenter is explicitly described and requested. Effects in studies such as mine, where participants would also need to formulate beliefs about the hypotheses that the experimenters are interested in—and independently decide that conforming to these hypotheses would constitute a favor to the experimenter—are likely to be yet smaller. The fact that no participant mentions anything resembling experimenter demand effects in their description of how they made choices (see Section 4.5.2) further argues against this concern.

That said, I of course cannot fully rule out that participants in my study might have managed to anticipate the hypotheses being tested, (spuriously) adjusted their behavior to an unprecedentedly large extent, and avoided mentioning this at the end of the experiment.

Risk Aversion and Caution

Next, a reasonable concern is that, if participants are uncertain about the value of various attributes, some results might be driven by small-stakes risk-aversion or caution à la Cerreia-Vioglio et al. (2015). For example, if participants cannot remember exactly how much money a red box or a nickel is worth, risk aversion or caution might lead them both to react less to its true value and to shy away from options that include more subjective uncertainty. Several results point against this interpretation. First, the value of some attributes is straightforward (pennies, nickels, dimes). Second, all my results hold even among the correct-beliefs sample who accurately reports the value of all attributes.[14] Third, and most tellingly, notice that if uncertainty (plus either risk aversion or caution) were driving the main effects, we would expect that information about the target attribute should continue to boost responsiveness to it even when information about a new attribute is provided in Period 3 (Treatment 4): presumably, uncertainty about the target attribute would be eliminated after 20 choices where its value is prominently displayed. Instead we see responsiveness to the target attribute revert to its baseline level, consistent with an attention interpretation.

6 Discussion: Implications for Models of Attention

Most of the focus so far has been on the implications of my findings for information interventions and preference estimation. But many of my results speak to recent models in behavioral economics of how attention is allocated. Most existing theories in economics model attention as a function solely of objective payoffs. For example, several recent theories posit a role for the range of payoffs across attributes within a given choice. We saw in Section 4.2 some evidence for such a channel: participants pay somewhat less attention to attributes when they take on larger values, consistent with relative thinking (Bushong et al. 2021) or diminishing sensitivity (Bordalo et al. 2013) though not with focusing (Koszegi & Szeidl 2013). But these forces appeared modest compared to the other effects in the experiment.

In addition, few of the results described in Section 4 are predicted by off-the-shelf models of rational inattention. First, the fact that inattention takes the form of neglect challenges the contention that, conditional on what agents attend to, they processes information rationally. Second, the workhorse model of rational inattention (see Mackowiak et al. 2022 for a review) posits that Shannon entropy governs the costs of attention. An implication of this assumption

[14]Of course, strictly speaking participants could be uncertain about attribute values but nonetheless manage to correctly guess their values in my beliefs questions. Even in this case, we should expect results to be *weaker* for the correct-beliefs sample, under the intuitive assumption that people who report correct beliefs are less uncertain than those who report incorrect beliefs. This is not what I find.

is that the relative probability of different choices should respond only to differences in their objective payoffs, which is inconsistent with large differences I find in responsiveness at baseline to equal-value changes across attributes.[15] Gabaix (2014) assumes that agents allocate their limited attention toward attributes that are more important for their choice. Consistent with this idea, participants reacted most to Option A's certain value, which had the largest variance and therefore was most often pivotal. On the other hand, we saw significant differences in attention to Option B's coins compared to its colored boxes, despite (by design) extremely similar distributions of payoffs across these two types of attributes and therefore no difference in their importance. Next, the sensitivity of attention to obviously redundant information is not predicted by any of these models, unless they are added as separate assumptions about the cognitive costs of accessing information (which I discuss more below).

Instead, many of my experimental results point toward the need to incorporate contextual factors as important determinants of attention. An example of such a model is salience theory (Bordalo et al. 2022), in which limited attention is drawn bottom-up to features of the environment that are prominent or surprising. Two aspects of my results resonate with this framework. First, and most obviously, the fact that clearly irrelevant information ("12 pennies are worth \$0.12!") has such large effects on attention highlights the importance of the immediate choice environment and which things are made visually prominent within it. Baseline attention toward the attributes in my experiment also corroborate this channel: the feature drawing the most attention (Option A's certain value) was also the one displayed separately and in larger font, while the attribute that most escaped attention (Option A's lottery) was the only one not explicitly depicted on every decision screen.[16]

Second, past contexts and choices have significant lasting effects on attention. For example, having directed attention toward an attribute in the past by providing information continues to dramatically boost attention toward that attribute in the future (so long as no new information about a different attribute takes its place). More speculatively, the

[15] More precisely, the following equation should hold:

$$\ln P(B|\omega_1) - \ln P(B|\omega_2) = \frac{1}{k}\Big[u(B|\omega_1) - u(B|\omega_2)\Big]$$

where ω_1 and ω_2 are two states (i.e., realizations of the attributes of Options A and B), $P(B|\omega)$ is the probability that Option B is chosen given ω, $u(B|\omega)$ is the payoff from choosing B given ω, and k is the cost of attention. By letting ω_1 and ω_2 be states that differ only along one attribute, we can see that equal-value changes in any attribute should yield identical differences in choice probabilities. See Dean & Neligh (2023).

[16] The importance of visual prominence also brings to mind theories like the attentional drift-diffusion model (e.g., Yang & Krajbich 2022), where gaze modulates cognitive accumulation of evidence from different aspects of the decision. Some of my results, like those on distraction and neglect, are consistent with this framework. However, these theories tend not to model the agent's decision of where to look, and so to that extent they cannot explain why attention shifts in the way it does in my experiment.

fact that coins (which participants have much experience attending to) draw more attention than colored boxes (which they do not) is consistent with a similar channel. Additionally, attributes draw attention in part when they are surprising, in that they take on different values than what the agent has experienced in the past, as we saw by comparing attention toward attributes that change vs are frozen throughout the experiment.[17]

One natural way to reconcile some of my results with models in the spirit of the rational inattention literature would be to postulate an attention-cost function that depends on more than Shannon entropy or objective stakes (e.g., see Hébert & Woodford 2021). For example, perhaps certain attributes are simply cognitively cheaper to assess (e.g., depending on how they are visually displayed, whether information is describing them, or whether similar types of information have been considered in the past). Many participants, after all, were aware that they were only imperfectly attempting to choose the highest-value option (see Section 4.5.2), an attitude very much in the spirit of costly information processing models and one not typically emphasized in so-called "bottom-up" theories of attention. I view incorporating (perhaps salience-driven) contextual factors into such models as providing a way of combining attractive elements of both bottom-up and top-down theories of attention. Further, even if some of the results of my experiment are influenced by non-payoff-based attention costs, this would not undermine any of the implications these results have for the design and interpretation of information interventions.

7 Conclusion

In this paper, I show experimentally that inattention can badly distort naive revealed preference estimates. Information can both exacerbate and ameliorate this problem by shifting attention. In particular, information shifts attention toward the attribute it describes while distracting away from other attributes. Further, the "default" value—how agents treat attributes to the extent that they fail to pay attention to them—appears to be less than their prior/beliefs about them, suggesting that inattention takes the form of neglect. All these attentional effects appear despite the information in my experiment often being transparently redundant.

Next, I find that attention is fragile. Information starkly shifts what agents attend to, but new information can quickly undo such effects. This suggests that to be effective attention-boosting interventions (e.g., to improve choices or estimate preferences) need to operate close

[17] In addition to giving a key role to prominence and the role of past experiences, salience theory also typically assumes that attention toward one feature is decreasing in the salience of other features (distraction) and that the default value absent attention is zero (neglect).

to the moment where a relevant decision is being made. An open question is what factors contribute to the longevity of attention effects, and how these forces interact with similar fragility in belief updating (e.g., see Graeber et al. 2024).

Both passive and active control-group information experiments in general lead to biased estimates of agents preferences, though the latter at least have the correct sign. I introduce a novel paradigm called parallel active control-group experiments (PACEs), which yield unbiased preferences estimates if information-induced attention across attributes is equal. This assumption will tend not to be testable in field settings, but it appears satisfied in my controlled experimental environment. Thus, PACEs represent a potentially useful tool to economists hoping to leverage information experiments to learn about welfare-relevant preferences.

References

Alesina, A., Miano, A., & Stantcheva, S. (2023, 1). Immigration and redistribution. *The Review of Economic Studies*, *90*, 1-39. doi: 10.1093/restud/rdac011

Allcott, H., & Taubinsky, D. (2015). Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *American Economic Review*, *105*(8), 2501–2538. doi: 10.1257/aer.20131564

Altmann, S., Grunewald, A., & Radbruch, J. (2021). Interventions and cognitive spillovers. *The Review of Economic Studies*. doi: 10.1093/restud/rdab087

Barrera, O., Guriev, S., Henry, E., & Zhuravskaya, E. (2020, 2). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of Public Economics*, *182*. doi: 10.1016/j.jpubeco.2019.104123

Bolte, L., & Raymond, C. (2025). *Emotional inattention.*

Bordalo, P., Burro, G., Coffman, K., Gennaioli, N., & Shleifer, A. (2025). Imagining the future: memory, simulation, and beliefs. *Review of Economic Studies*, *92*(3), 1532–1563.

Bordalo, P., Conlon, J. J., Gennaioli, N., Kwon, S. Y., & Shleifer, A. (2025). How people use statistics. *Review of Economic Studies*.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2013). Salience and Consumer Choice. *Journal of Political Economy*, *121*(5), 803–843. doi: 10.1086/673885

Bordalo, P., Gennaioli, N., & Shleifer, A. (2020). Memory, attention, and choice. *Quarterly Journal of Economics*.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2022). Salience. *Annual Review of Economics*, 1-42.

Bradley, S., & Feldman, N. E. (2020). Hidden baggage: Behavioral responses to changes in airline ticket tax disclosure. *American Economic Journal: Economic Policy*, *12*, 58-87. doi: 10.1257/pol.20190200

Brown, J., Hossain, T., & Morgan, J. (2010). Shrouded Attributes and Information Suppression: Evidence from the Field. *The Quarterly Journal of Economics*, *125*(2), 859–876.

Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in saudi arabia. *American economic review*, *110*(10), 2997–3029.

Bushong, B., Rabin, M., & Schwartzstein, J. (2021). A model of relative thinking. *Review of Economic Studies*.

Cerreia-Vioglio, S., Dillenberger, D., & Ortoleva, P. (2015). Cautious expected utility and the certainty effect. *Econometrica*, *83*, 693-728. doi: 10.3982/ecta11733

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, *99*, 1145-1177. doi: 10.1257/aer.99.4.1145

Cohn, A., & Maréchal, M. A. (2016, 12). Priming in economics. *Current Opinion in Psychology*, *12*, 17-21. doi: 10.1016/j.copsyc.2016.04.019

Colonnelli, E., Gormsen, N. J., & McQuade, T. (2024). Selfish corporations. *Review of Economic Studies*, *91*(3), 1498–1536.

Conlon, J. J. (2021). Major Malfunction: A Field Experiment Correcting Undergraduates' Beliefs about Salaries. *Journal of Human Resources*. doi: 10.2139/ssrn.2965743

Conlon, J. J., & Patel, D. (2025). What jobs come to mind? stereotypes about fields of study.

Cullen, Z., & Perez-Truglia, R. (2022). How much does your boss make? the effects of salary comparisons. *Journal of Political Economy*, *130*(3), 766–822.

Dai, W., Yang, T., White, B. X., Palmer, R., sanders, E. K., McDonald, J. A., ... Albarracín, D. (2023). Priming behavior: A meta-analysis of the effects of behavioral and nonbehavioral primes on overt behavioral outcomes. *Psychological Bulletin*. doi: 10.1037/bul0000374.supp

Dean, M., & Neligh, N. (2023). Experimental tests of rational inattention. *Journal of Political Economy*.

De Quidt, J., Haushofer, J., & Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review*, *108*(11), 3266–3302.

Enke, B. (2020). What you see is all there is. *The Quarterly Journal of Economics*, *135*(3), 1363–1398.

Enke, B., & Zimmermann, F. (2019). Correlation neglect in belief formation. *The review of economic studies*, *86*(1), 313–332.

Falk, A., & Zimmermann, F. (2024). Attention and dread: Experimental evidence on preferences for information. *Management Science*, *70*(10), 7090–7100.

Gabaix, X. (2014). A Sparsity-Based Model of Bounded Rationality. *The Quarterly Journal of Economics*, 1661–1710. doi: 10.1093/qje/qju024.Advance

Gabaix, X. (2019). *Behavioral inattention* (Vol. 2). Elsevier B.V. doi: 10.1016/bs.hesbe.2018.11.001

Gagnon-Bartsch, T., Rabin, M., & Schwartzstein, J. (2023). *Channeled attention and stable errors.*

Golman, R., & Loewenstein, G. (2018). Information gaps: A theory of preferences regarding the presence and absence of information. *Decision*, *5*(3), 143.

Graeber, T. (2023). Inattentive inference. *Journal of the European Economic Association*, *21*(2), 560–592.

Graeber, T., Roth, C., & Zimmermann, F. (2024). Stories, statistics, and memory. *The Quarterly Journal of Economics*, *139*(4), 2181–2225.

Haaland, I., & Roth, C. (2020). Labor market concerns and support for immigration. *Journal of Public Economics*, *191*, 104256.

Haaland, I., Roth, C., & Wohlfart, J. (2023, 3). Designing information provision experiments. *Journal of Economic Literature*, *61*, 3-40. doi: 10.1257/jel.20211658

Hanna, R., Mullainathan, S., & Schwartzstein, J. (2014). Learning through noticing: Theory and evidence from a field experiment. *The Quarterly Journal of Economics*, *129*(3), 1311–1353.

Hébert, B., & Woodford, M. (2021, 10). Neighborhood-based information costs. *American Economic Review*, *111*, 3225-3255. doi: 10.1257/AER.20200154

Imai, T., Pace, D. D., Schwardmann, P., & van der Weele, J. J. (2024). Correcting consumer misperceptions about co2 emissions.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, *125*(2), 515–548.

Koszegi, B., & Szeidl, A. (2013). A Model of Focusing in Economic Choice. *Quarterly Journal of Economics*, 53–104. doi: 10.1093/qje/qjs049.Advance

Mackowiak, B., Matejka, F., & Wiederholt, M. (2022). Rational inattention: A review. *Annual Review of Economics*. doi: 10.2866/417246

Roth, C., Schwardmann, P., & Tripodi, E. (2024). Depression stigma.

Schulze-Tilling, A. (2025). The effectiveness of carbon labels.

Schwartzstein, J. (2014). Selective attention and learning. *Journal of the European Economic Association*, *12*, 1423-1452. doi: 10.1111/jeea.12104

Stantcheva, S. (2023). How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics*, *15*(1), 205–234.

Taubinsky, D., & Rees-Jones, A. (2018). Attention variation and welfare: Theory and evidence from a tax salience experiment. *Review of Economic Studies*, *85*(4), 2462–2496. doi: 10.1093/restud/rdx069

Winichakul, K. P., Lezema, G., Mustafi, P., Lepper, M., Wilson, A., Danz, D., & Vesterlund, L. (2024). The effect of experimenter demand on inference.

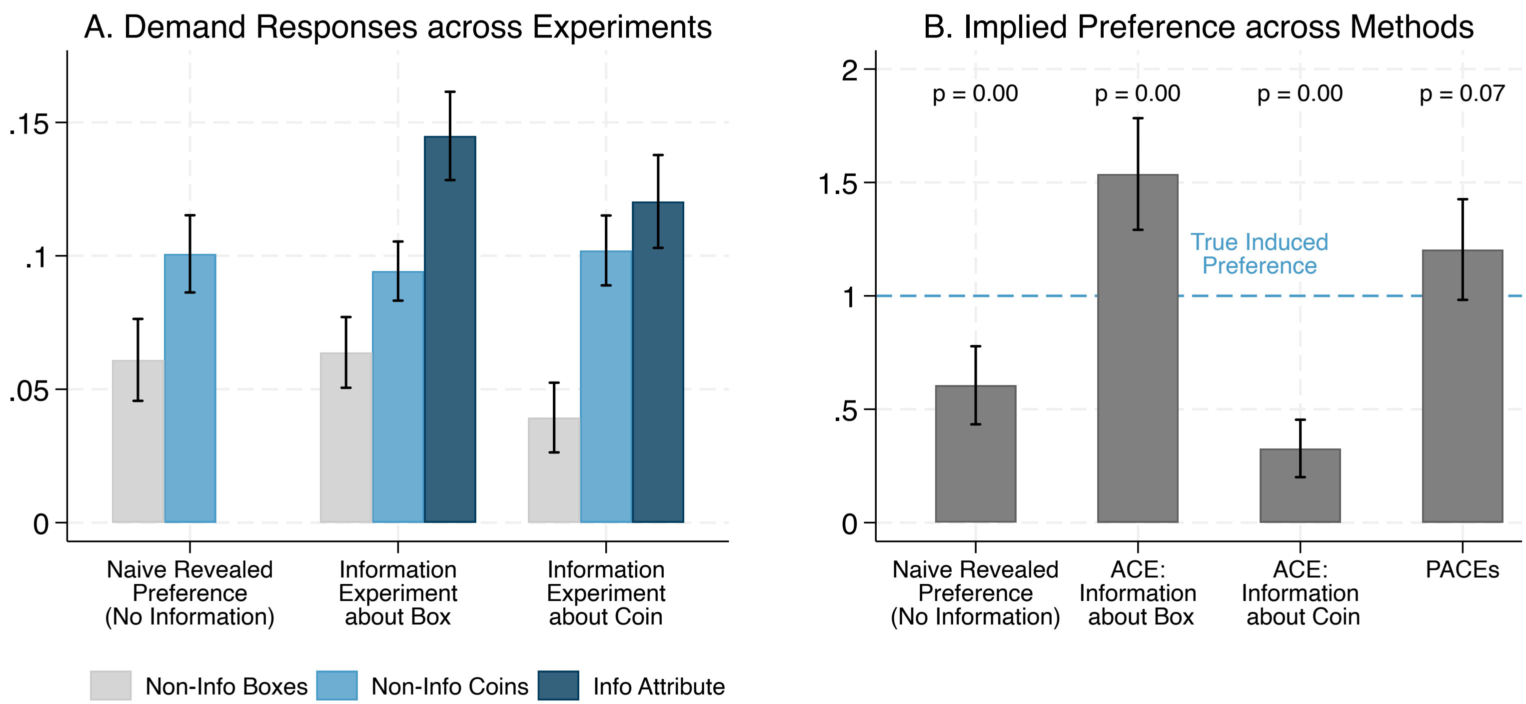
Wiswall, M., & Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, *82*, 791-824.

Wojtowicz, Z., Molnar, A., Golman, R., & Loewenstein, G. (2025). Willful inattention: Keeping aversive information out of mind. *Current Opinion in Psychology*, 102116.

Yang, X., & Krajbich, I. (2022). A dynamic computational model of gaze and choice in multi-attribute decisions. *Psychological Review*.

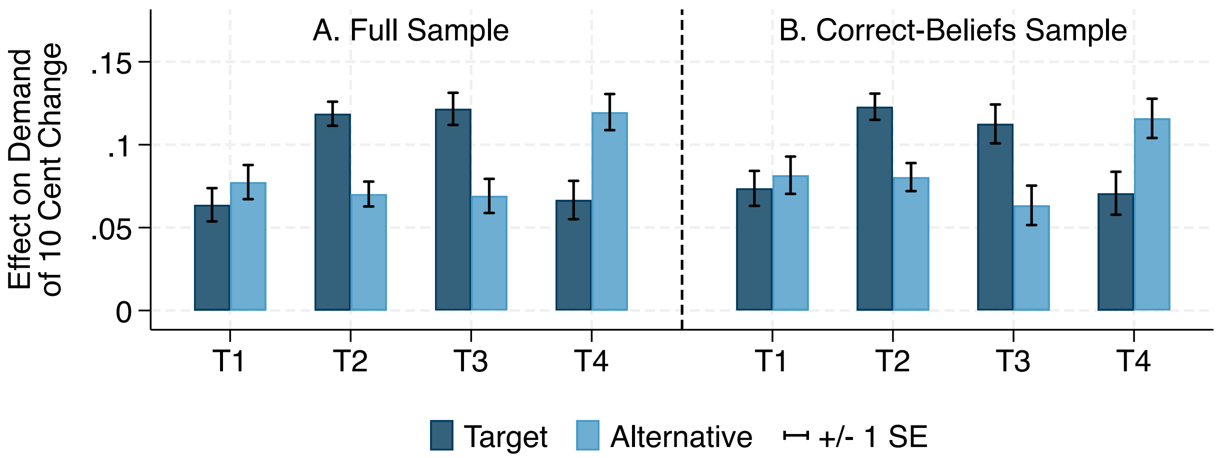
A For Online Publication: Additional Figures and Tables

Figure A.I: Recovering Preferences: Correct-Beliefs Sample



Notes: Panel A of this figure summarizes OLS estimates of equation 10 (see columns 4-6 of Table A.V for the full estimates) among the correct-beliefs sample. The first pair of bars shows estimates of α_1 and α_2—responsiveness of demand for Option B to the value of its coins and colored boxes—among participants in Period 2 who received no information (treatment 1). The next three bars in Panel A show estimates of α_1, α_2, and α_3 for participants (in treatments 2-4) who received information about a coin in Period 2. The final three bars in Panel A show analogous estimates for participants who received information abotu a colored box in Period 2. Panel B then computes preference estimates assuming different methods. The first bar shows estimated preferences for blocks vs colored boxes assuming "naive revealed preference": i.e., dividing the estimate for α_1 by α_2 among those who received no information in Period 2 (gray and light blue bars in leftmost pair of bars in Panel A). The next bar shows the quotient of α_1 and α_3 among those receiving information about a coin (dark blue divided by light blue from middle three bars in Panel A). The next bar shows the quotient of α_3 and α_2 among those receiving information about a block (gray divided by dark blue from right three bars in Panel A). The final bar in Panel B shows estimates from the parallel active control group experiments (PACEs) method: dividing α_3 among those receiving information about a box (dark blue from middle bars in Panel A) by α_3 among those receiving information about a coin (dark blue from rightmost bars in Panel A). Whiskers show 95% confidence intervals from robust standard errors clustered at the individual level.

Figure A.II: Dynamics of Information Effects



Notes: This figure depicts a subset of the OLS estimates from Table A.VI, which estimates equation 9 using the Period 3 decisions of each treatment group (separately). The dependent variable is whether the participant chose Option B. The independent variables are the certain value of Option A, the target attribute of Option B, the alternative attribute of Option B, and the sum of the four other attributes of Option B. Treatment 1 (T1) never received information about the target attribute. Treatments 2-4 (T2-T4) received information about the target attribute in Period 2, but differed in the information presented during Period 3. During this period, Treatment 2 received no information, Treatment 3 continued to receive information about the target attribute, and Treatment 4 received information about the alternative attribute (randomly chosen from the remaining four non-frozen attributes of Option B). Panel A shows estimates including all participants, while Panel B shows estimates including only participants who correctly respond to unincentivized questions at the end of the experiment about the monetary value of each coin and every possible colored box. Whiskers show robust standard errors, clustered at the individual level. This figure shows only the coefficients on the target attribute (dark blue bars) and alternative attribute (light blue bars). Table A.VI shows the full regression results for these specifications.

Table A.I: Attention at Baseline

	Full Sample			Correct-Beliefs Sample			Rational Choice		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Option A Value	0.128***	0.143***	0.128***	0.126***	0.131***	0.127***	0.129***	0.059***	0.129***
	(0.002)	(0.007)	(0.002)	(0.002)	(0.008)	(0.002)	(0.002)	(0.006)	(0.002)
Option A Lottery	0.014***	0.022**	0.014***	0.016***	0.018	0.017***	0.124***	0.123***	0.124***
	(0.003)	(0.011)	(0.003)	(0.004)	(0.012)	(0.004)	(0.005)	(0.005)	(0.005)
Option B Coins	0.096***	0.126***		0.098***	0.122***		0.129***	0.144***	
	(0.004)	(0.013)		(0.004)	(0.015)		(0.002)	(0.007)	
Option B Boxes	0.045***	0.052***		0.051***	0.073***		0.128***	0.151***	
	(0.004)	(0.013)		(0.004)	(0.014)		(0.002)	(0.007)	
Option A Value^2		0.002**			0.001			-0.008***	
		(0.001)			(0.001)			(0.001)	
Option A Lottery^2		0.001			0.000				
		(0.001)			(0.001)				
Option B Coins^2		-0.012**			-0.010			-0.006**	
		(0.005)			(0.006)			(0.003)	
Option B Boxes^2		-0.004			-0.011			-0.012***	
		(0.006)			(0.007)			(0.003)	
Option B Changing			0.073***			0.079***			0.128***
			(0.002)			(0.003)			(0.002)
Option B Frozen			0.054***			0.049***			0.131***
			(0.011)			(0.012)			(0.003)
Constant	0.621***	0.621***	0.643***	0.619***	0.592***	0.641***	0.503***	0.362***	0.503***
	(0.023)	(0.030)	(0.023)	(0.027)	(0.035)	(0.027)	(0.013)	(0.017)	(0.013)
Observations	11,800	11,800	11,800	8,740	8,740	8,740	11,800	11,800	11,800
Individuals	590	590	590	437	437	437	590	590	590
R^2	0.35	0.35	0.34	0.35	0.35	0.34	0.60	0.61	0.60
p-value: Option A Value = Boxes	0.00	0.00		0.00	0.00		0.71	0.00	
p-value: Option A Lottery = Boxes	0.00	0.06		0.00	0.00		0.41	0.00	
p-value: Coins = Boxes	0.00	0.00		0.00	0.02		0.78	0.48	
p-value: Squared terms all zero		0.01			0.22			0.00	
p-value: Moving = Frozen			0.08			0.02			0.38

Notes: Each column shows OLS estimates of modifications of equation 9 using the Period 1 decisions of all treatment groups. The dependent variable is whether the participant chose Option B. The independent variables in columns 1, 4, and 7 include the certain value of Option A, the subjective value of Option A's lottery, the sum of Option B's coins, and the sum of Option B's colored boxes. The subjective value of the lottery is calculated by multiplying the prize for winning the lottery with each participants' prior belief about their odds of winning (winsorized at the 90th percentile). The certain and lottery values of Option A are multiplied by negative one so that the expected sign of all uninteracted attribute coefficients is positive. Columns 2, 5, and 8 additionally include the squared values of these attributes (squaring the individual components and then taking the sum). The independent variables in columns 3, 6, and 9 are identical to columns 1 and 4 for Option A, but for Option B include, first, the sum of the five changing attributes and, second, the attribute that was frozen throughout the experiment at its initial value. Columns 1-3 include all participants, columns 4-6 include only participants who correctly respond to unincentivized questions at the end of the survey about the monetary value of each coin and every possible colored box, and columns 7-9 replace participants' actual choices with the expected-payoff maximizing decision. Robust standard errors, clustered at the individual level, are presented in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.II: Attention at Baseline: Logit Specification

	Full Sample			Correct-Beliefs Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
Option A Value	0.792***	1.099***	0.779***	0.806***	1.122***	0.796***
	(0.025)	(0.073)	(0.025)	(0.027)	(0.080)	(0.024)
Option A Lottery	0.089***	0.137*	0.089***	0.108***	0.115	0.108***
	(0.021)	(0.073)	(0.021)	(0.027)	(0.077)	(0.024)
Option B Coins	0.605***	0.796***		0.633***	0.785***	
	(0.030)	(0.099)		(0.030)	(0.080)	
Option B Boxes	0.286***	0.341***		0.336***	0.472***	
	(0.025)	(0.080)		(0.030)	(0.103)	
Option A Value^2		0.036***			0.037***	
		(0.007)			(0.008)	
Option A Lottery^2		0.005			0.001	
		(0.007)			(0.008)	
Option B Coins^2		-0.077**			-0.060*	
		(0.037)			(0.032)	
Option B Boxes^2		-0.027			-0.067	
		(0.038)			(0.050)	
Option B Changing			0.456***			0.509***
			(0.020)			(0.024)
Option B Frozen			0.340***			0.319***
			(0.059)			(0.077)
Constant	0.685***	1.094***	0.797***	0.682***	1.061***	0.797***
	(0.138)	(0.216)	(0.153)	(0.166)	(0.222)	(0.176)
Observations	11,800	11,800	11,800	8,740	8,740	8,740
Individuals	590	590	590	437	437	437
p-value: Option A Value = Boxes	0.00	0.00		0.00	0.00	
p-value: Option A Lottery = Boxes	0.00	0.07		0.00	0.00	
p-value: Coins = Boxes	0.00	0.00		0.00	0.02	
p-value: Squared terms all zero		0.00			0.00	
p-value: Moving = Frozen			0.05			0.01

Notes: Each column shows logit estimates of modifications of equation 9 using the Period 1 decisions of all treatment groups. The dependent variable is whether the participant chose Option B. The independent variables in columns 1 and 4 include the certain value of Option A, the subjective value of Option A's lottery, the sum of Option B's coins, and the sum of Option B's colored boxes. The subjective value of the lottery is calculated by multiplying the prize for winning the lottery with each participants' prior belief about their odds of winning (winsorized at the 90th percentile). The certain and lottery values of Option A are multiplied by negative one so that the expected sign of all uninteracted attribute coefficients is positive. Columns 2 and 5 additionally include the squared values of these attributes (squaring the individual components and then taking the sum). The independent variables in columns 3 and 6 are identical to columns 1 and 4 for Option A, but for Option B include, first, the sum of the five changing attributes and, second, the attribute that was frozen throughout the experiment at its initial value. Columns 1-3 include all participants, while columns 4-6 include only participants who correctly respond to unincentivized questions at the end of the survey about the monetary value of each coin and every possible colored box. Robust standard errors, clustered at the individual level, are presented in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.III: Attention at Baseline for Treatment 1 Correct-Beliefs Sample

	OLS			Logit		
	(1)	(2)	(3)	(4)	(5)	(6)
Option A Value	0.124***	0.108***	0.124***	0.764***	0.931***	0.760***
	(0.006)	(0.018)	(0.006)	(0.065)	(0.217)	(0.057)
Option A Lottery	0.024***	-0.025	0.025***	0.151***	-0.173	0.156**
	(0.009)	(0.026)	(0.009)	(0.058)	(0.195)	(0.062)
Option B Coins	0.092***	0.116***		0.580***	0.734***	
	(0.008)	(0.034)		(0.071)	(0.210)	
Option B Boxes	0.055***	0.079***		0.359***	0.498***	
	(0.008)	(0.025)		(0.065)	(0.156)	
Option A Value^2		-0.002			0.018	
		(0.002)			(0.021)	
Option A Lottery^2		-0.005*			-0.035	
		(0.003)			(0.022)	
Option B Coins^2		-0.010			-0.060	
		(0.014)			(0.080)	
Option B Boxes^2		-0.012			-0.066	
		(0.011)			(0.069)	
Option B Changing			0.075***			0.478***
			(0.006)			(0.050)
Option B Frozen			0.063**			0.397**
			(0.026)			(0.171)
Constant	0.666***	0.541***	0.685***	0.902***	0.567	1.002***
	(0.052)	(0.067)	(0.053)	(0.334)	(0.591)	(0.332)
Observations	2,020	2,020	2,020	2,020	2,020	2,020
Individuals	101	101	101	101	101	101
R^2	0.33	0.34	0.33			
p-value: Option A Value = Boxes	0.00	0.37		0.00	0.10	
p-value: Option A Lottery = Boxes	0.02	0.00		0.01	0.00	
p-value: Coins = Boxes	0.00	0.39		0.00	0.34	
p-value: Squared terms all zero		0.19			0.08	
p-value: Moving = Frozen			0.67			0.63

Notes: Each column shows OLS (columns 1-3) or logit (columns 4-6) estimates of modifications of equation 9 using participants' Period 1 decisions with the following modifications. The dependent variable is whether the participant chose Option B. The independent variables in columns 1-2 and 4-5 include the certain value of Option A, the subjective value of Option A's lottery, the sum of Option B's coins, and the sum of Option B's colored boxes. The subjective value of the lottery is calculated by multiplying the prize for winning the lottery with each participants' prior belief about their odds of winning (winsorized at the 90th percentile). Columns 2 and 5 include the squared values of these attributes (squaring the individual components and then taking the sum). The certain and lottery values of Option A are multiplied by negative one so that the expected sign of all uninteracted attribute coefficients is positive. The attributes in columns 3 and 6 are identical to columns 1 and 4 for Option A, but for Option B include, first, the sum of the five changing attributes and, second, the attribute that was frozen throughout the experiment at its initial value. Data are restricted to Treatment-1 participants (who do not receive any information about Option-B attributes throughout the experiment) who correctly respond to unincentivized questions at the end of the survey about the monetary value of each coin and every possible colored box. Robust standard errors, clustered at the individual level, are presented in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.IV: Effects of Information about the Target Attribute

	Full Sample			Correct-Beliefs Sample		
	Pooled (1)	Coin (2)	Box (3)	Pooled (4)	Coin (5)	Box (6)
Info	0.022***	0.013	0.041***	0.022***	0.014	0.039***
	(0.007)	(0.010)	(0.010)	(0.008)	(0.011)	(0.011)
Option A Value X No Info	0.127***	0.132***	0.123***	0.125***	0.131***	0.121***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Option A Value X Info	0.129***	0.133***	0.125***	0.128***	0.134***	0.124***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Option A Lottery X Info	0.000	-0.000	0.001	-0.001	0.000	-0.002
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Target Attribute X No Info	0.074***	0.101***	0.047***	0.074***	0.097***	0.053***
	(0.005)	(0.006)	(0.006)	(0.005)	(0.007)	(0.007)
Target Attribute X Info	0.125***	0.122***	0.132***	0.132***	0.122***	0.143***
	(0.006)	(0.008)	(0.008)	(0.006)	(0.009)	(0.008)
Other Changing Option B Attributes X No Info	0.075***	0.070***	0.080***	0.080***	0.073***	0.087***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Other Changing Option B Attributes X Info	0.069***	0.063***	0.074***	0.073***	0.062***	0.082***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.005)	(0.004)
Frozen Option B Attribute X Info	-0.017*	-0.030**	-0.003	-0.014	-0.023*	-0.004
	(0.009)	(0.012)	(0.012)	(0.009)	(0.013)	(0.012)
Observations	23,600	11,400	12,200	17,480	8,200	9,280
Individuals	590	285	305	437	205	232
R^2	0.513	0.517	0.512	0.523	0.526	0.525
Effect on Target	0.051	0.022	0.085	0.057	0.025	0.089
p-value: Effect on Target is Zero	0.000	0.020	0.000	0.000	0.021	0.000
Total Effect on Non-Targets	-0.038	-0.056	-0.025	-0.040	-0.065	-0.021
p-value: Total Effect on Non-Targets is Zero	0.019	0.011	0.272	0.020	0.009	0.376
Total Effect on All Attributes	0.014	-0.034	0.059	0.017	-0.040	0.069
p-value: Total Effect on All Attributes is Zero	0.430	0.141	0.023	0.375	0.127	0.013

Notes: This table shows OLS estimates of modifications of equation 9 using the Periods 1 and 2 decisions of all treatment groups. The dependent variable is whether the participant chose Option B. The independent variables are certain value of Option A, the subjective value of Option A's lottery, the target attribute of Option B, the sum of the four other changing attributes of Option B, and the frozen Option B attribute, and individual-fixed effects. The certain and lottery values of Option A are multiplied by negative one so that the expected effect of increasing all attributes is positive. "Info" is a dummy variable for receiving information about the target attribute (i.e., being in Treatments 2-4 during Period 2). Columns 1-3 include all participants, while columns 4-6 include only participants who correctly respond to unincentivized questions at the end of the survey about the monetary value of each coin and every possible colored box. The main effect of the Option A Lottery and Frozen Option B Attribute are excluded because there are co-linear with the individual-fixed effects (and interactions of these attributes with information represent the difference in responsiveness when receiving information compared to not). The row showing the "Effect on Target" (and associated p-value) refers to the difference between the coefficient on "Target Attribute" with and without information. The row showing the "Effect on Non-Targets" (and associated p-value) first adds the coefficients for all attributes other than the target attribute (multiplying the coefficient on the other changing Option B attributes by four since there were four such attributes) and takes the difference between this value with and without information. "Total Effect on All Attributes" is analogous but also includes the target attribute. Robust standard errors, clustered at the individual level, are presented in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.V: Recovering Preferences

	Full Sample			Correct-Beliefs Sample		
	No Info	Info about Box	Info about Coin	No Info	Info about Box	Info about Coin
	(1)	(2)	(3)	(4)	(5)	(6)
Target Value		0.133^{***}	0.121^{***}		0.145^{***}	0.120^{***}
		(0.008)	(0.008)		(0.008)	(0.009)
Option A Value	0.128^{***}	0.125^{***}	0.133^{***}	0.123^{***}	0.123^{***}	0.135^{***}
	(0.005)	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)
Non-Info Option B Boxes	0.054^{***}	0.048^{***}	0.037^{***}	0.061^{***}	0.064^{***}	0.039^{***}
	(0.007)	(0.006)	(0.006)	(0.008)	(0.007)	(0.007)
Non-Info Option B Coins	0.102^{***}	0.091^{***}	0.107^{***}	0.101^{***}	0.094^{***}	0.102^{***}
	(0.006)	(0.005)	(0.006)	(0.007)	(0.006)	(0.007)
Observations	2,680	4,700	4,420	2,020	3,560	3,160
Individuals	134	235	221	101	178	158

Notes: This table shows OLS estimates of modifications of equation 9 using the Periods 2 decisions of Treatment 1 participants (columns 1 and 4), treatments 2-4 participants whose target attribute is a colored box (columns 2 and 5), or participants whose target attribute is a coin (columns 3 and 6). The dependent variable is whether the participant chose Option B. The independent variables are certain value of Option A, the target attribute of Option B (in columns 2, 3, 5, and 6), the sum of coins that information is not describing, the sum of boxes that information is not describing, and individual-fixed effects. The certain value of Option A is multiplied by negative one so that the expected effect of increasing all attributes is positive. Columns 1-3 include all participants, while columns 4-6 include only participants who correctly respond to unincentivized questions at the end of the survey about the monetary value of each coin and every possible colored box. The Option A Lottery is excluded because it is co-linear with the individual-fixed effects. Robust standard errors, clustered at the individual level, are presented in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.VI: Dynamics of Information Effects

	Full Sample				Correct-Beliefs Sample			
	T1	T2	T3	T4	T1	T2	T3	T4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Target Option B Attribute	0.064***	0.119***	0.122***	0.067***	0.074***	0.123***	0.113***	0.071***
	(0.010)	(0.007)	(0.010)	(0.012)	(0.011)	(0.008)	(0.012)	(0.013)
Alternative Option B Attribute	0.077***	0.070***	0.069***	0.120***	0.082***	0.080***	0.063***	0.116***
	(0.010)	(0.007)	(0.010)	(0.011)	(0.011)	(0.008)	(0.012)	(0.012)
Other Option B Attributes	0.068***	0.075***	0.070***	0.059***	0.073***	0.080***	0.077***	0.073***
	(0.006)	(0.004)	(0.007)	(0.007)	(0.007)	(0.005)	(0.007)	(0.007)
Option A Value	0.132***	0.131***	0.127***	0.134***	0.130***	0.132***	0.124***	0.128***
	(0.004)	(0.003)	(0.004)	(0.005)	(0.005)	(0.003)	(0.005)	(0.006)
Observations	2,680	4,780	2,280	2,060	2,020	3,580	1,600	1,540
Individuals	134	239	114	103	101	179	80	77
R^2	0.56	0.55	0.57	0.55	0.57	0.56	0.57	0.55
p-value: Target Same as T1		0.00	0.00	0.85		0.00	0.01	0.86
p-value: Target Same as T2			0.81	0.00			0.46	0.00
p-value: Target Same as T3				0.00				0.02
p-value: Alternative Same as T1		0.57	0.57	0.00		0.94	0.27	0.04

Notes: This table shows OLS estimates of modifications of equation 9 using the Period 3 decisions of each treatment group (separately by column, treatments indicated in the column headings). The dependent variable is whether the participant chose Option B. The independent variables are the certain value of Option A, the target attribute of Option B, the alternative attribute of Option B, the sum of the four other attributes of Option B, and individual-fixed effects. The certain value of Option A is multiplied by negative one, such that the expected sign of all coefficients is positive. Treatment 1 (T1) never received information about the target attribute. Treatments 2-4 (T2-T4) received information about the target attribute in Period 2, but differed in the information presented during Period 3. During this period, Treatment 2 received no information, Treatment 3 continued to receive information about the target attribute, and Treatment 4 received information about the alternative attribute. Columns 1-4 include the full sample, while columns 5-8 include only participants who correctly respond to unincentivized questions at the end of the survey about the monetary value of each coin and every possible colored box. Robust standard errors, clustered at the individual level, are presented in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

B For Online Publication: Proofs

Here I provide proofs of the results in Section 2. Note that, in general, for any states s and s':

$$\begin{aligned}
\Delta_{s\to s'} D &= \frac{1}{\kappa}\Big(u(s) - u(s')\Big) \\
&= \frac{1}{\kappa}\sum_{k=1}^{K} v_k \left[\theta_k(s')\cdot \widehat{a_k}(s') + \big(1-\theta_k(s')\big)\cdot \bar{a}_k - \theta_k(s)\cdot \widehat{a_k}(s) - \big(1-\theta_k(s)\big)\cdot \bar{a}_k\right] \\
&= \frac{1}{\kappa}\sum_{k=1}^{K} v_k \left[\theta_k(s')\cdot \Delta_{s\to s'}\hat{a}_k + \Delta_{s\to s'}\theta_k\Big(\hat{a}_k(s) - \bar{a}_k\Big)\right] \qquad (11)
\end{aligned}$$

Then,

$$\begin{aligned}
\Delta_{s\to s_k^+} D &= \frac{1}{\kappa} v_k \theta_k(s_k^+)\Delta_{s\to s_k^+}\hat{a}_k \\
\Delta_{s\to s_j^+} D &= \frac{1}{\kappa} v_j \theta_k(s_j^+)\Delta_{s\to s_j^+}\hat{a}_j \\
\implies \frac{\Delta_{s\to s_k^+} D}{\Delta_{s\to s_j^+} D} &= \frac{v_k}{v_j}\frac{\Delta_{s\to s_k^+}\hat{a}_k}{\Delta_{s\to s_j^+}\hat{a}_j}\frac{\theta_k(s)}{\theta_j(s)}
\end{aligned}$$

because by assumption attention is unchangedL $\theta(s) = \theta(s_j^+) = \theta(s_k^+)$. This proves Proposition 1.

Next, for Proposition 2, we assumed beliefs about j are always correct. Therefore:

$$\begin{aligned}
\Delta_{i(s)\to i(s_k^+)} D &= \frac{1}{\kappa} v_k \theta_k^I \\
\implies \frac{\Delta_{i(s)\to i(s_k^+)} D}{\Delta_{s\to s_j^+} D} &= \frac{v_k}{v_j}\frac{\theta_k^I}{\theta_j^S} \\
\implies \frac{\Delta_{i(s)\to i(s_k^+)} D}{\Delta_{s\to s_j^+} D} - \frac{\Delta_{s\to s_k^+} D}{\Delta_{s\to s_j^+} D} &= \frac{v_k}{v_j}\frac{1}{\theta_j^S}\left[\theta_k^I - \Delta_{s\to s_k^+}\hat{a}_k\theta_k^S\right]
\end{aligned}$$

which, with some rearranging, is Proposition 2.

Next, note that Proposition 3 is just the expectation of equation 11 (given the assumption that attention is constant across states within I_k and within S.

Finally, Proposition 4 follows straightforwardly from equation 11 by substituting in the correct terms.

C For Online Publication: Data Appendix

Here I provide additional details on the experiment. Participants were recruited through Prolific to participate in a "Quick Survey on Decision Making." Potential survey-takers were not told anything of the content of the survey except that it was part of a research study and that it would be more difficult to complete if they struggled to tell colors apart. A total of 590 participants completed the survey during December 2022. All participants are US residents, the average participant is 41 years old, and 51% are women. The median respondent took 21 minutes to complete the experiment. The experiment paid a $3.00 completion fee plus any bonus that participants earned from their choices. The full experimental survey can be viewed by following this link: https://harvard.az1.qualtrics.com/jfe/form/SV_5jbhR2VLKqq2n6S.

Participants had to correctly answer a series of comprehension checks in order to continue with the survey. If they initially answered a question incorrectly, they were forced to revise their answer before proceeding. I do not exclude anyone from the data for poor performance, but 95% of comprehension questions were answered correctly on the first try, suggesting a high level of engagement and understanding. Table B.I gives more details on these comprehension questions and provides a link to the experimental instructions.

In Section 4.5.2, I describe open-ended responses that participants gave describing how they made their decisions. These come from a question at the very end of the experiment (i.e., after all choices and belief elicitation) asking "In your own words, how did you choose between option A and option B? Did you use any strategies to try to simplify the problem?". An open text box appeared below the question where participants could (though they were not required to) write a response. I then feed these responses into a large language model (GPT-4o) to derive binary yes-no responses to the following questions about participants' responses (rates of "Yes" responses in parentheses):

1. "Does this respondent mention attempting to add up the value of the options, or choosing the option that seemed to add up to the highest amount?" (66.8%)
2. "Does this respondent mention trying to identify or guessing or approximating what the best option was (rather than knowing what it was)?" (42.5%)
3. "Does this respondent mention anything reacting to reminders or information (other than the baseline rules of the experiment)?" (1.2%)
4. "Does this respondent mention anything about what the researcher's hypothesis might have been (or altering their decisions at all based on what the researcher seemed to want)?" (0.0%)
5. "Does this respondent mention anything about doing what the experiment *wanted*

or *suggested to* her to do (or any variation thereof), rather than what she otherwise thought was the right choice?" (0.0%)

Two differences between the preregistration and the analysis that appears in the main text bear mentioning. First, the preregistration mentions that some participants see information in period 2 about the alternative attribute and no information in period 3. However, because participants are not told which are the target and alternative attributes, this treatment is equivalent to being told about the target attribute (i.e., simply changing labels for which are the "target" and "aternative" attributes). I combine these into Treatment group 2 after this relabeling.

Second, the preregistration mentions a second experiment ($N = 211$) with the following differences from the experiment in the main text. First, there was no lottery associated with Option A. Second, the level of attributes was chosen such that the target attribute of Option B was always pivotal in deciding which option yielded the higher bonus. That is, whenever the pivotal attribute took on either its intermediate or high value (recall that all Option B attributes had three possible values), Option B was the payoff-maximizing choice. The other "non-pivotal" attributes were by construction uncorrelated with the payoff-maximizing decision.

Just like in the main experiment, no one saw information in the first period. Unlike in the main experiment, in each of the remaining three periods, participants either saw no information, information about the (pivotal) target attribute, or information about a randomly selected non-target (and therefore non-pivotal) attribute. This randomization occurred across periods and within participant.

This experiment was intended to test how the welfare effect of information depends on whether it directs attention toward pivotal or non-pivotal attributes. Table B.II shows an OLS regression where the dependent variable is an indicator for whether the participant chose the lower-value option. It regresses this variable on individual-fixed effects and indicators for whether the participant was receiving information about the pivotal target attribute or a non-pivotal non-target attribute. We see that, while both types of information reduce the rate at which participants mistakenly choose the lower-value option, these effects are larger when the information is about a pivotal attribute ($p < 0.01$). One explanation for why even information about non-pivotal attributes may have improved decision-making is that in Period 1 (i.e., absent any information) participants only choose Option B 41% of the time despite it being the payoff-maximizing choice 70% of the time. Thus, because paying more attention to the non-pivotal attribute boosts demand for Option B, it therefore also reduces mistakes. These results are again consistent with the default value being less than participants' priors.

Table B.I: Comprehension Questions

	Topic	% Correct on First Attempt
Question #1	One choice is randomly chosen to be implemented	97.1%
Question #2	How bonus is calculated if Option A is chosen	87.4%
Question #3	Value of coins in Option B	98.3%
Question #4	Value of each color for top box	93.9%
Question #5	Value of each color for middle box	95.6%
Question #6	Value of each color for bottom box	96.5%
Question #7	Information is provided randomly	Not recorded*

Notes: The instructions that participants saw, along with the text of each comprehension question can be found at **this link**. *The fraction of participants who correctly answered question #7 on the first attempt was not recorded due to a coding error, so this statistic is unavailable.

Table B.II: Effects of Information about Pivotal and Non-Pivotal Information

	Mistake (1)
Information about Pivotal Attribute	-0.083***
	(0.011)
Information about Non-Pivotal Attribute	-0.046***
	(0.012)
Observations	16,880
Individuals	211
R^2	0.23
p-value: Information Effects Equal	0.00

Notes: This table shows an OLS regression where the dependent variable is an indicator for whether the participant chose the lower-value option. It regresses this variable on individual-fixed effects and indicators for whether the participant was receiving information about the pivotal target attribute and non-pivotal non-target attribute. Robust standard errors, clustered at the individual level, are presented in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.