

# Temptation to Consume Information\*

Vivek Roy-Chowdhury<sup>†</sup>

June 2023

## Abstract

We provide evidence that information is intrinsically *tempting*: it can be costly to resist because doing so requires self-control. By eliciting menu preferences over information, we find that around half of individuals are tempted by information they do not want to see. Some participants are exogenously offered information in a later session, regardless of their menu preferences. Analysis of choices and deliberation times indicates self-control costs are economically significant and intrinsic information preferences are dynamically inconsistent. Since the availability of temptations can harm welfare, our findings provide novel insight into the implications of recent growth in the supply of information.

---

\*Pre-registered on the AEA RCT Registry at <https://doi.org/10.1257/rct.10865-2.0>. This project was supported by the Keynes Fund and the Monica Kornberg Memorial Fund at the University of Cambridge. I am very grateful to Andrew Caplin, Dorothea Kuebler, Elliot Lipnowski, Kirby Nielsen, Itzhak Rasooly, Christopher Rauh, Julia Shvets, Sarah Taylor, Severine Toussaert, Florian Zimmermann, and attendees of various presentations for their valuable comments.

<sup>†</sup>Faculty of Economics, University of Cambridge; [vr277@cam.ac.uk](mailto:vr277@cam.ac.uk)

# 1 Introduction

The global capacity for information processing grew at a compound annual growth rate of 58% from 1986–2007 (Hilbert and López, 2011), and more recent innovations in computing and AI have introduced unprecedented changes in the instantaneous availability of complex information (Bubeck et al., 2023). This staggering expansion in supply suggests that decisions to avoid or consume information have become an increasingly important determinant of economic welfare. A major contribution of recent research has been to establish that information is avoided in several contexts, especially when it leads to undesirable choices or creates unpleasant feelings.<sup>1</sup> Nonetheless, casual observations and emerging evidence abound that information is still sought in many such situations, even when it has no direct relevance for decision making.

This paper investigates whether individuals are *tempted* by information they would rather not see, meaning that avoiding it is costly to welfare. To do so, we experimentally measure whether individuals have temptation and self-control preferences (Gul and Pesendorfer, 2001; Toussaert, 2018) over information. Critically, the information we offer does not have any decision-making (instrumental) value. Otherwise, observed preferences for information could conceal preferences for related choices. In our case, participants discover that they will be offered information on whether a prior decision to claim a bonus payment resulted in a charity donation being reduced.

We obtain three main findings. First, almost half of individuals have strict menu preferences implying they are tempted by information they do not want. Second, a small share of individuals tempted by information make dynamically inconsistent choices, and self-control costs are behaviourally meaningful and economically significant for the remainder. Third, unavoidable partial information erodes the value of commitments to avoiding information.

Our main empirical exercise involves revealing temptation through ex ante preferences for commitment. However, commitment can be motivated by either self-control costs or dynamically inconsistent choices. Recent empirical work on temptation essentially infers a role for costly self-control because individuals strictly prefer commitments even when they do not make dynamically inconsistent choices (Toussaert, 2018). A novel feature of our design is that we use recorded deliberation time under temptation to quantify self-control costs. Strikingly, we find that those willing to pay more for commitment devices have a harder time turning information down when it is offered to them a day later, exactly as predicted by models of temptation with sophistication. This suggests intrinsic preferences for information may often be dynamically inconsistent even if choices are not, with notable implications for

---

<sup>1</sup>See Golman et al. (2017) for an overview.

ongoing empirical and theoretical work.

Partial information is seemingly deployed to increase the demand for information in many contexts — for instance, notifications on mobile devices and social media. Using a treatment condition in which an initial signal is automatically shown in the second session, we find evidence that unfavourable partial information weakens one’s *ex ante* resolve to avoid further information. Our suggested interpretation is that partial unfavourable information restricts an individual’s ability to engage in wishful thinking about the consequences of the selfish action. More broadly, our findings also introduce an argument for policies treating *ex ante* undesirable information like other well-known temptation goods: sometimes, it may be better for such information not to be manufactured at all. We speculate that this could be an important dimension of policy assessments of some recent developments in social media and consumer access to artificial intelligence.

**Related literature.** In early theoretical work by Kreps and Porteus (1978), ‘intrinsic’ preferences for information are the mirror image of preferences over the temporal resolution of uncertainty. In their model and subsequent extensions (Epstein and Zin, 1989; Grant et al., 1998), individuals can prefer to expedite or delay the receipt of information on the outcome of a lottery depending on properties of their risk and time preferences. In a more significant departure from the canonical expected utility model, a substantial body of recent work suggests individuals could avoid information because of its role in anticipatory feelings. This model was pioneered in Caplin and Leahy (2001) has been extended in several directions (Caplin and Leahy, 2004; Brunnermeier and Parker, 2005; Epstein, 2008; Kőszegi and Rabin, 2009; Ely et al., 2015). Dynamic inconsistency in intrinsic preferences for information could be a natural extension of such theories, and our results indicate that such efforts may be warranted. Separately, our results also suggest that in settings where information theoretically has instrumental value, it may be valuable to model both intrinsic and instrumental costs of avoiding information rather than just the latter.

Empirical evidence on intrinsic information preferences has overwhelmingly focused on exploring whether individuals prefer to resolve uncertainty sooner rather than later (Eliaz and Schotter, 2007; Ganguly and Tasoff, 2017; Nielsen, 2020; Falk and Zimmermann, 2023). Recent work demonstrates that individuals are often impatient to acquire even potentially discomfiting information (Masatlioglu et al., 2022). Separately, there is growing empirical evidence that information with possible instrumental value is avoided or incorrectly recalled in a wide variety of settings, including in the field (Eil & Rao, 2011; Oster et al., 2013; Zimmermann, 2020; Huffman et al., 2022; Roy-Chowdhury, 2022; Saccardo & Serra-Garcia, 2023).

While a standard interpretation of existing empirical results is that information is acquired only when it has instrumental or affective benefits at the margin, our results suggest that acquiring unpleasant information could additionally be determined by individuals' capacities for self-control. If temptation is an important determinant of information consumption, existing results are compatible with both information-averse and information-loving preferences: information consumption may reflect failed self-control rather than straightforward welfare optimisation. To emphasise this point, we reiterate our result that individuals who seem to have a harder time resisting instantaneous offers of information are willing to pay *more* for commitments to avoiding it in advance.

We also contribute to the broader empirical literature on temptation and self-control. The main innovation of the Gul and Pesendorfer (2001) model is that the presence of tempting options in the choice set may be costly even when they are not consumed. This prediction is supported by many empirical studies of procrastination and effort (Alan and Ertac, 2015; Royer et al., 2015; Toussaert, 2018; Sadoff et al., 2020). The main objective of our paper is to bring research on temptation to the novel domain of preferences for information. However, we also highlight a methodological innovation in our experimental measurement of self-control costs through deliberation time: our paper is the first in the literature on temptation and self-control to provide an empirical quantification of self-control costs and whether they appear to motivate commitment demand. As previously highlighted, our data are basically unequivocal in their support for the Gul and Pesendorfer (2001) model, in which agents are sophisticated about their future self-control costs.

The paper proceeds as follows. In Section 2, we outline a model of temptation and self-control, based closely on Gul and Pesendorfer (2001), which generates the key set of empirical tests connecting menu preferences to costly self-control. In Section 3, we provide details of the the experiment and sample. Section 4 contains the results of the experiment. Therein, Section 4.1 considers menu preferences, Sections 4.2 and 4.3 examines information choices and self-control costs in session 2, Section 4.4 compares preferences across our treatment conditions, and Section 4.5 explores some qualitative measures of preferences. In Section 5, we discuss the implications of our results for theoretical and empirical research on information preferences, as well as policy.

## 2 Temptation and self-control

Our goal in this section is to apply an established model of temptation and self-control (Gul and Pesendorfer, 2001) to intrinsic preferences for information.<sup>2</sup> This exercise generates the two key empirical tests underpinning our experiment. The first is whether individuals strictly prefer not to be *offered* undesired information at a future date, revealing that it is a temptation. The second is whether menu preferences reflect sophistication about the strength of temptation, either through dynamically inconsistent choice or (observed) costs of self-control.

Consuming the information is given by the action  $a = 1$ , while not consuming it is given by  $a = 0$ . Key to the experiment is that all participants submit preference rankings over all possible menus for  $a$  in session 1, a day before the information is made available. That is, they provide preference rankings over the set of all three possible choice menus for  $a$ :  $\{\{0\}, \{0, 1\}, \{1\}\}$ . To be clear,  $\{0\}$  involves committing to not seeing the information,  $\{0, 1\}$  involves being given the choice to see the information, and  $\{1\}$  involves committing to see the information. As we will see shortly, participants' preferences over these three menus (prior to the period of choice) reveal a great deal about whether they are *tempted* by information, as well as their motives for commitment.

From Gul and Pesendorfer (2001), a temptation makes the decision maker strictly worse off if chosen, but it also makes them strictly worse off if it is present in the choice set.

**Definition 1:** *Information is a temptation if  $\{0\} \succ_1 \{0, 1\}$  and  $\{0\} \succ_1 \{1\}$ .*

In the first session of our experiment, which uses the menu preference design from Toussaert (2018), individuals rank all three menus,  $\{0\}$ ,  $\{0, 1\}$ , and  $\{1\}$ . Thus, by recording menu preferences across the sample, we immediately reveal how many individuals are tempted by the information we offer. However, our experiment is designed to say more than just how many individuals are tempted by information. There are two menu preferences satisfying Definition 1:  $\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$  and  $\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$ . It turns out that selection between these two menu preferences, as well as individuals' willingness to pay for commitments, is behaviourally meaningful. In order to elaborate on these points, we set out an explicit representation of temptation preferences closely following Gul and Pesendorfer (2001).<sup>3</sup>

---

<sup>2</sup>We do not set out an explicit model for information preferences at this stage; our experiment aims to test whether resisting information requires self-control, rather than to distinguish between models of information preferences. However, we present a simple model which can rationalise many of our results in Appendix A3.

<sup>3</sup>Other models of temptation, self-control, and present bias include O'Donoghue and Rabin (1999), Fudenberg and Levine (2006), Noor (2007), Brocas and Carrillo (2008), Heidhues and Köszegi (2009), and Fudenberg and Levine (2012). These models can be seen as extensions of the general framework set out in Gul and Pesendorfer (2001), and focus on features of preferences which are ancillary to our experiment. Most

**Model.** Gul and Pesendorfer (2001) provide two representations supporting Definition 1. The first is the ‘self-control’ type, which does not succumb to temptation but suffers a self-control cost from the presence of tempting options in the choice set. The second is the ‘overwhelming temptation’ type, which succumbs to temptation. Somewhat in the vein of Dekel and Lipman (2012), we relax the assumption that preferences are deterministic; we allow individuals to be unsure of which type they will be in period 2.

We consider a 2-period setting. In period 2, the decision maker chooses whether to obtain information by setting the action  $a$  from the menu  $\mathcal{M} \in \{\{0\}, \{0, 1\}, \{1\}\}$ . As noted,  $a = 0$  avoids information while  $a = 1$  consumes it. Key to the self-control problem is that the decision maker has two (conflicting) evaluations of acquiring the information through setting  $a = 1$ . The first is their commitment utility,  $g_1(a)$ , capturing their evaluation of information without any temptation. The second is their temptation utility,  $g_2(a)$ , the evaluation of information under the full strength of temptation.

The individual believes that with probability  $\tilde{p}$ , they will be the ‘self-control’ type. As in Gul and Pesendorfer (2001), this type chooses  $a$  to maximise  $g_1(a) + g_2(a)$ , and so engages in a compromise between their conflicting evaluations of  $a$ . However, in doing so, they face a self-control cost through the term  $-g_2(a')$ , where  $a'$  is the optimal action for  $g_2(a)$  in  $\mathcal{M}$ . With probability  $1 - \tilde{p}$ , they will be the ‘overwhelming temptation’ type: they are unable to resist temptation and they simply select  $a'$ . The individual evaluates menus  $\mathcal{M} \in \{\{0\}, \{0, 1\}, \{1\}\}$  as below:

$$V(\mathcal{M}) = \tilde{p} \left( \max_{a \in \mathcal{M}} [g_1(a) + g_2(a)] - g_2(a') \right) + (1 - \tilde{p}) g_1(a'), \text{ s.t. } g_2(a') > g_2(z) \text{ for all } z \neq a' \in \mathcal{M}. \quad (1)$$

Suppose that  $a = 0$  is the optimal action for  $g_1(a) + g_2(a)$  within  $\mathcal{M} = \{0, 1\}$ , so that the individual avoids information in period 2 if she turns out to be the ‘self-control’ type. Information is *tempting* when  $a = 1$  is the optimal choice in  $\mathcal{M} = \{0, 1\}$  for temptation utility  $g_2$ . Under these conditions, the representation above implies a strict preference for avoiding information in period 1,  $\{0\} \succ_1 \{1\}$ , as well as commitment,  $\{0\} \succ_1 \{0, 1\}$ . The former is immediately visible from (1); implementing  $\{1\}$  relative to  $\{0\}$  results in a guaranteed payoff of  $g_1(1)$  rather than  $g_1(0)$ , and we know  $g_1(a)$  must be maximised at  $a = 0$  from the fact that  $g_1(a) + g_2(a)$  is maximised at  $a = 0$  and  $g_2(a)$  is maximised at  $a = 1$ . To see that  $\{0\} \succ_1 \{0, 1\}$ , consider the marginal value of  $\{0\}$  relative to  $\{0, 1\}$ :

$$V(\{0\}) - V(\{0, 1\}) = \tilde{p} [-g_2(0) + g_2(1)] + (1 - \tilde{p}) [g_1(0) - g_1(1)] > 0. \quad (2)$$

---

commonly, they generate richer dynamics of temptation, self-control costs, and sophistication, which are irrelevant to our simple two-period setting.

The first term in (2) captures the marginal benefit of commitment for the tempted type, and the second captures the same for the dynamic inconsistency type. For the ‘self-control’ type,  $\{0\}$  eliminates the cost of self-control relative to  $\{0, 1\}$ : temptation utility is trivially maximised at  $a = 0$  when the choice set is  $\{0\}$ . On the other hand, the ‘overwhelming temptation’ type is aware that  $a = 1$  will be chosen if  $\{0, 1\}$  is implemented. The benefit corresponding to this type is therefore the marginal benefit of  $a = 0$  under commitment utility: commitment actually changes the choice made in period 2. Thus, temptation encompasses two possibilities: that undesirable information is consumed (dynamically inconsistent choice); and that not consuming it is costly (costly self-control).

Aside from permitting type uncertainty, another modelling detail we need to introduce relative to Gul and Pesendorfer (2001) is a cost of submitting each strict menu preference: participants in our experiment must complete a short task to individually confirm each of the (at most 2) strict preferences they submit. Usefully, this means that  $\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$  captures more than just the limiting case where  $\tilde{p} = 0$ .

**Proposition 1:** *When submitting strict preferences is costly, for menu preferences  $V(\mathcal{M})$  where  $g_1(0) + g_2(0) > g_1(1) + g_2(1)$  and  $g_2(1) > g_2(0)$ ,*

1. *Tempted types ( $\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$ ) expect small self-control costs,  $m_2$ , relative to their commitment preference for avoiding information,  $m_1$ . The subjective probability  $\tilde{p}$  of successful self-control increases the importance of  $m_2$ .*
2. *Strongly tempted types ( $\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$ ) either expect  $m_2 > m_1$ , or expect  $m_1 > m_2$  but with smaller  $\tilde{p}$  than tempted types.*

*Standard information averse types ( $\{0\} \sim_1 \{0, 1\} \succ_1 \{1\}$ ) expect small  $m_2$  and have small  $m_1$ . Higher  $\tilde{p}$  increases the importance of the first factor relative to the second.*

A corollary captures predictions on willingness to pay for menu preferences:

**Corollary 1:** *When submitting strict preferences is costly, for tempted types,*

1. *Willingness to pay for the preference  $\{0\} \succ_1 \{0, 1\}$  increases in self-control costs  $m_2$  and the commitment preference to avoid information  $m_1$ . A higher  $\tilde{p}$  increases the importance of  $m_2$  relative to  $m_1$  in determining willingness to pay.*
2. *Willingness to pay for the preference  $\{0, 1\} \succ_1 \{1\}$  decreases in self-control costs  $m_2$  but increases in the commitment preference to avoid information  $m_1$  and  $\tilde{p}$ .*

*Proof.* Suppose the utility cost of effort is  $c$ , and let the commitment value of avoiding information be  $m_1 = g_1(0) - g_1(1) > 0$  and the self-control cost be  $m_2 = g_2(1) - g_2(0) > 0$ . The expected marginal value of submitting  $\{0\} \succ_1 \{0, 1\}$  relative to  $\{0\} \sim_1 \{0, 1\}$  is

$$(1 - \tilde{p})m_1 + \tilde{p}m_2 - c, \quad (3)$$

while the expected marginal value of  $\{0, 1\} \succ_1 \{1\}$  relative to  $\{0, 1\} \sim_1 \{1\}$  is<sup>4</sup>

$$\tilde{p}(m_1 - m_2) - 4c. \quad (4)$$

Proof of the both the Proposition and the Corollary follows immediately. □

As above, we refer to the type  $\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$  as ‘strongly tempted’: either the temptation is likely to be irresistible, or self-control costs are high when it is resistible. Notably, a stronger preference for avoiding information under commitment utility, captured by a larger  $m_1$ , increases the likelihood of satisfying both conditions and submitting the ‘tempted’ preference ranking  $\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$ . Thus, individuals of this type are likely to have the strongest preference for avoiding information, but still expect costs of self-control if offered it.

The parameter  $\tilde{p}$ , capturing the probability of successful self-control, plays a different role in the two conditions. In (3), it shapes the relative importance of the marginal commitment and marginal temptation payoffs to avoiding information. That is, if  $\tilde{p}$  is low and the individual feels they are likely to succumb to temptation, the period 1 value of avoiding information is more important than self-control costs in the decision to submit  $\{0, 1\} \succ_1 \{1\}$ . In (4), it is only relevant if  $m_1$  is larger than  $m_2$ , so if self-control costs are small compared to the commitment value of information. If so, a larger  $\tilde{p}$  — a higher probability of resisting temptation — makes it more likely that  $\{0, 1\} \succ_1 \{1\}$  is submitted. Intuitively, if self-control costs are high, being offered  $\{0, 1\}$  is unpleasant. The prospect of a struggle for restraint is more attractive if 0 is eventually chosen.

Finally, consider those submitting the menu preference  $\{0\} \sim_1 \{0, 1\} \succ_1 \{1\}$ . Without any cost of submitting strict preferences, this ranking would reflect that  $a = 0$  is the optimal action for both  $g_1(a)$  and  $g_2(a)$  within  $\{0, 1\}$ : the individual is not tempted by information. When taking costs into account, it is easy to establish using similar logic as before that

---

<sup>4</sup>This second equation involves a larger cost term because there is a 40% chance that first choices are implemented, a 10% chance that second choices are implemented, and a 0% chance that third choices are implemented.



some individuals who are tempted by information may now submit  $\{0\} \sim_1 \{0, 1\} \succ_1 \{1\}$ . Since in this case (4) is positive but (3) is not, these individuals are more likely to face low costs of self-control, have a low commitment value of avoiding information, and may perceive a very low probability of being dynamically inconsistent. Thus, even with effort costs in the elicitation procedure, interpretation of these types as not seeing themselves as strongly tempted by information is appropriate.

### 3 Experiment design and sample

Participants submit their preferences in session 1 over the feasible set of information menus they could face in session 2. Importantly, the information in our experiment is designed to be non-instrumental; it relates to the consequences of an unalterable choice. However, it could have affective value: individuals may feel guilty to discover that their choice negatively affected a charity donation.

The main objective of the experiment is to examine the frequency of the two menu preference rankings implying information is tempting:  $\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$  and  $\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$ . However, since many participants receive  $\{0, 1\}$  in session 2 regardless of their menu preferences, we can test whether menu preferences are explained by self-control costs or dynamic inconsistency in the way predicted by Proposition 1 and Corollary 1 by exploiting a naturally occurring measure of self-control costs.

#### 3.1 Experiment design

The menu preference exercise at the core of our experiment is based on Toussaert (2018). Her experiment tests for Gul and Pesendorfer (2001) temptation in the context of supplying effort during a tedious task. In her experiment, subjects anticipate completing a tedious task. The tempting option is to sacrifice earnings by reducing effort and reading a distracting story. She collects participants' menu preferences prior to the task and finds that about one-quarter have what we refer to as the 'tempted' preference  $\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$ .

Our experiment applies the same exercise to non-instrumental information. As previously outlined, our empirical tests have two major components. The first uses menu preferences (from a first session) to revealing whether individuals are tempted by information. The second uses data on choices and behaviour (from a second session) under randomly implemented menus to test whether menu preferences reflect sophistication about temptation, consistent with Proposition 1 and Corollary 1.

**Session 1.** Subjects complete a short task involving moving sliders to specified values. They are then informed they can claim a bonus payment of \$4 or forgo it. The base payment for session 1 is \$1.40, meaning accepting the \$4 substantially increases participants’ pay. The only benefit of forgoing the bonus is that a ‘minority’ of participants (in reality, 15%) have been allocated to be *Donors*, such that taking the bonus results in a charity donation being reduced by \$15. The remainder of participants are not *Donors*, meaning there is no consequence of taking the \$4.

After choosing whether to claim the bonus, participants are told the name of the charity, and are informed that they will be able to find out if they were a *Donor* in session 2. The information is contained in a virtual envelope, which is generated in session 2 unless the choice menu  $\{0\}$  is implemented. Participants then submit preferences over what information access options they will have in session 2. To ensure strict preferences for information in session 2, opening the envelope costs money. The price is randomly selected from  $\{\$0.25, \$0.50, \$0.75, \$1\}$ , and participants are informed of the price they will eventually face when submitting menu preferences in session 1.

**Session 2.** In this session, which takes place one or two days after session 1, participants choose whether to open the envelope, depending on the choice menu implemented. The menu they face is influenced by their preference ranking from session 1, although as we mention below, half of participants face the menu  $\{0, 1\}$  regardless of their preferences.

**Menu preferences and willingness to pay.** Towards the end of session 1, participants submit preference rankings over the *menus* of information that will eventually be offered to them in session 2. To aid comprehension, the session 2 information choice is framed as opening a virtual envelope, and commitment is framed as the envelope not being generated for the participant. The set of menus in session 1 then has the following presentation:

- Automatically open the envelope. ( $\{1\}$ )
- Do not generate the envelope at all. ( $\{0\}$ )
- Ask me if I want to open the envelope next time. ( $\{0, 1\}$ )

Importantly, participants are informed that session 2 will take ‘about the same amount of time’ regardless of which menu is implemented.<sup>5</sup> To ensure full comprehension and incentive compatibility, we follow a multi-step elicitation procedure in which participants must repeatedly confirm any strict preferences they submit. The corresponding experiment

---

<sup>5</sup>In practice, this is implemented using buffer tasks when participants receive  $\{0\}$  in session 2.

pages can be viewed in the Appendix, starting with Figure A3. Initially, participants give each menu a numerical rank between 1 and 3; multiple menus can be given the same rank. They then navigate to a second page where the menus are displayed in a re-randomised order and they cannot proceed until their preference rankings on the two pages match. Then, where any strict preferences have been submitted, participants proceed to the second stage of the procedure.

It is at this second stage that we construct our main measure of menu preferences. Participants are prompted to confirm that they ‘definitely have a clear preference’ between each adjacent pair of menus in their ranking, and are advised that they will have to complete a short task to confirm their preference. This means our primary outcome measure is somewhat conservative, as participants who are not willing to pay the effort cost to confirm each strict preference are defined as being indifferent. The task is an exact repetition of the one they completed at the start of the experiment, involving dragging sliders to numerical values, and participants are able to amend their preferences upon seeing the task.

After the effort task, we elicit participants’ willingness to pay to maintain the first strict preference in their final ranking using the Becker-DeGroot-Marschak (BDM) mechanism. Participants are told that a random payment between \$0 and \$0.50 will be generated. If that number is higher than their stated willingness to pay, which is also constrained to be between \$0 and \$0.50, they receive the payment and their first and second choices are swapped. Response values are restricted to multiples of \$0.05, starting from \$0.

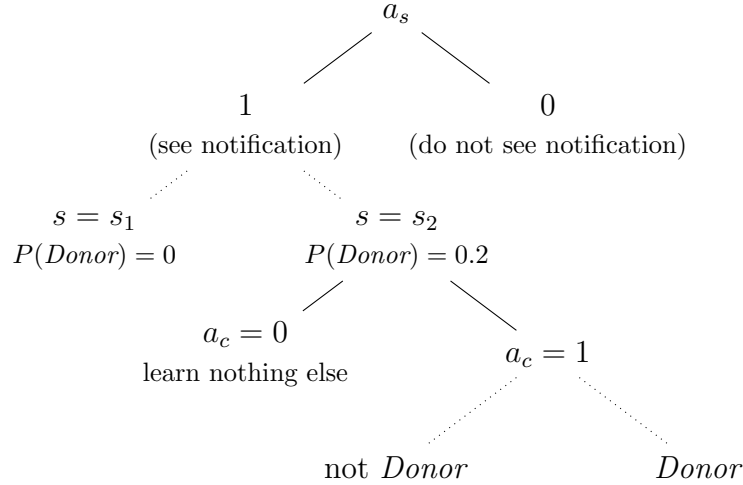
After the result of the BDM auction, menu preferences are implemented in session 2 according to the following regime, which ensures incentive compatibility: with 50% probability, the menu  $\{0, 1\}$  is exogenously implemented so participants must decide whether to access information regardless of their preferences. Otherwise, menu preferences are relevant: with 40% probability, first choices are implemented; with 10% probability, second choices are implemented. If participants are indifferent between any two options, one of them is randomly selected to be strictly preferred and the scheme is implemented as usual.

**Treatment conditions.** As previously mentioned, we included two treatment conditions to investigate how anticipating unavoidable signals affects the attractiveness of commitment devices to avoiding information. Participants taking the bonus face one of two subtly different information offers in session 2.<sup>6</sup> In *FullChoice*, the envelope contains an initial signal  $s$  of whether or not the individual was a *Donor*. In particular,  $s$  can take values  $s_1$  or  $s_2$ , implying either a 0% or a 20% chance respectively that taking the \$4 bonus resulted in the charity

---

<sup>6</sup>The 9% of participants who turned down the bonus were all allocated to a condition in which opening the envelope immediately confirms their *Donor* status.

Figure 1: Choice sequence without commitment, for participants who take the bonus



Note:  $s$  is the signal.  $a_s$  is the action to see the signal.  $a_c$  is the action to confirm *Donor* status conditional on  $s = s_2$ .

donation being reduced by \$15. Conditional on  $s = s_2$ , participants can then choose to costlessly confirm whether the charity donation was reduced by \$15.

In *PartialChoice*, participants know they will automatically learn the value of  $s$  at the start of session 2. Instead, the envelope in session 2 is generated conditional on  $s = s_2$ . As illustrated by Figure 1, the key point is that ‘not generating the envelope’ in *PartialChoice* involves a participant committing to not learn whether they are a *Donor* when they already know there is a 20% chance they are one. In *FullChoice*, the same action means learning nothing about the chance of being a *Donor*.

### 3.2 Sample

The sample recruited from Prolific initially included 673 adults residing in the US and fluent in English. Of those, 669 were invited to session 2 after passing one of two attention checks in session 1, and 634 completed both sessions. Table 1 provides a summary of key statistics. We captured a rich range of ages within the sample, with a standard deviation of 12.3 years. Nonetheless, the median age was a little lower than the average in the US population. White respondents were overrepresented, and the employment rate (for the subset with recent data) was lower than in the general population. The sample was balanced on sex.

Numerous measures suggest the experimental data were of high quality. Only 11 of 673 participants failed either of the two attention checks inserted in session 1; 4 of those were immediately disqualified because they failed both attention checks. We exclude all 11 from the ensuing analysis. Moreover, 69% of participants chose to answer optional free-text questions in session 1 rationalising their submitted menu preferences. Finally, as already noted, 95% of

Table 1: Summary statistics

	Value	...of $N$
Mean age	36.7	662
Standard deviation of age	12.3	662
Male	50%	662
Ethnicity: white	72.5%	661
Ethnicity: black	5.9%	661
Ethnicity: mixed/other	11.4%	661
Mean altruism (0-10)	7.0	632
Mean risk (0-10)	4.7	632
Mean patience (0-10)	6.6	632
Mean self-control (0-10)	5.6	632
Took bonus	91.0%	669
Treatment: <i>FullAvoid</i>	47.9%	609
Attended session 2	94.8%	669
Employed	45.1%	478
Mean reward	\$5.80	634
Session 1 median duration	7m, 40s	669
Session 2 median duration	2m, 4s	634
Failed attention checks	11	673

*Note:* Qualitative, self-reported measures elicited at the end of session 2. Treatment allocation was random only for participants who took the bonus. Reported only for participants who completed both sessions. The few participants who did not complete session 2 were just paid the base pay for session 1, \$1.40.

participants recruited for session 1 returned for session 2 a day later. Compensation rates were relatively generous: the mean reward for the experiment was \$5.80, implying an average hourly pay rate of around \$34.80.

91% of participants accepted the bonus. Of the 9% who turned it down, 6 participants were allocated to be *Donors*. Accordingly, \$90 was donated to the charity after the experiment concluded, and all participants were contacted as promised after the experiment to inform them of the total donation amount (but nothing else).

## 4 Results

Our first set of results concern participants' menu preferences from session 1, which reveal whether information is tempting (Proposition 1). To support the interpretation that commitment is driven by temptation, we then ask whether there is any evidence of the two

motives for commitment outlined in Section 2: dynamically inconsistent choice, and costly self-control. For this purpose, we analyse whether participants access information, or appear to suffer self-control costs in resisting information, when their commitments are exogenously not implemented, and test the model’s precise predictions on heterogeneity in the strength of temptation across individuals. We then compare preferences across our treatment conditions, investigating whether anticipating partial information makes further information more or less tempting *ex ante*. Finally, we offer some qualitative evidence supporting the interpretation that information is tempting.

## 4.1 Menu preferences from session 1

**Result 1:** *49% of individuals do not want the information we offer but are tempted by it.*

A menu preference ranking with both  $\{0\} \succ_1 \{1\}$  and  $\{0\} \succ_1 \{0, 1\}$  implies information is seen as tempting: it encompasses a strict preference both to avoid information and to remove it from the later choice set, relative to being offered it. As detailed in Proposition 1, there are two types for whom this holds:  $\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$ , the ‘tempted’ type, and  $\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$ , the ‘strongly tempted’ type. However, 13 menu preference rankings are possible in total.

Table 2: Menu preferences from session 1

Menu preference	Type	N	Share
$\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$	Tempted	254	38.4% (1.9)
$\{0\} \sim_1 \{0, 1\} \sim_1 \{1\}$	Indifferent	80	12.1% (1.3)
$\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$	Strongly tempted	71	10.7% (1.2)
$\{0\} \sim_1 \{0, 1\} \succ_1 \{1\}$	Standard info. averse	54	8.2% (1.1)
$\{0, 1\} \succ_1 \{0\} \succ_1 \{1\}$	Flex	50	7.6% (1)
$\{0, 1\} \succ_1 \{1\} \succ_1 \{0\}$	Flex	34	5.1% (0.9)
$\{0, 1\} \succ_1 \{0\} \sim_1 \{1\}$	Flex	33	5% (0.8)
$\{1\} \succ_1 \{0, 1\} \succ_1 \{0\}$	Tempted (info. loving)	29	4.4% (0.8)
$\{0, 1\} \sim_1 \{1\} \succ_1 \{0\}$	Other	17	2.6% (0.6)
$\{0\} \succ_1 \{1\} \succ_1 \{0, 1\}$	Flexibility averse	14	2.1% (0.6)
$\{1\} \succ_1 \{0\} \sim_1 \{0, 1\}$	Other	13	2% (0.5)
$\{0\} \sim_1 \{1\} \succ_1 \{0, 1\}$	Commitment loving	9	1.4% (0.5)
$\{1\} \succ_1 \{0\} \succ_1 \{0, 1\}$	Flexibility averse	4	0.6% (0.3)
Total		662	100%

*Note:* Participants must complete a short task for each strict preference submitted in their ranking. If present, the first strict preference additionally requires participants to have a strictly positive willingness to pay to avoid a swap of their first and second choices. If  $WTP = 0$ , the first strict preference is replaced with indifference. Standard errors in parentheses, in percentage points.

The main results of the experiment are in Table 2, which reports the frequency of each full menu preference.<sup>7</sup> Using our primary measurement, information is tempting for 49% of individuals: not only do they strictly prefer not to see information ( $\{0\} \succ_1 \{1\}$ ), but they would explicitly rather not be *offered* information than be offered it ( $\{0\} \succ_1 \{0, 1\}$ ).<sup>8</sup> Within this set, the large majority (38% of the sample) are the ‘tempted’ type  $\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$ . Recalling Proposition 1, this type is tempted by information but has strong *ex ante* preferences to avoid information relative to self-control costs.

All 13 possible preference rankings are represented in the sample. This is even true in the first stage of elicitation, before participants are prompted to reconsider any strict preferences they submit (and pay a small effort cost to confirm them). However, the ‘tempted’ type is by far the most common of the 13 in the sample, followed by those who are indifferent between all three options, and then ‘strongly tempted’ types with the ranking  $\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$ . This last type also finds information tempting but expects either high self-control costs or a high chance of succumbing to temptation, such that the marginal value of being offered  $\{0, 1\}$  relative to just  $\{1\}$  is relatively small. As we will confirm when analysing session 2 behaviour, as well as our naturally occurring measure of self-control costs during session 2, costly self-control is an important motive for commitment.

Setting aside those who can be classified as finding information tempting, 18% of individuals in the sample have a strict preference for flexibility over either choice set restriction, suggesting uncertain preferences for information.<sup>9</sup> Within this set, binary preferences over  $\{0\}$  and  $\{1\}$  are approximately evenly distributed over the three possibilities.

Unsurprisingly, the prevalence of indifferences increases substantially from our initial ‘raw’ measurement of menu preferences to the main one, in which participants must complete a short effort task to confirm each strict preference (Table A2). To understand exactly what is going on here, note in the first elicitation stage, it is only weakly optimal for an individual who is indifferent between two options to submit an indifference. For example, take an individual whose true preference  $\succeq_1$  implies  $\{0, 1\} \sim_1 \{1\}$ . Submitting the preference  $\{0, 1\} \sim_1 \{1\}$  produces the same expected payoff as either  $\{0, 1\} \succ_1 \{1\}$  or  $\{1\} \succ_1 \{0, 1\}$ . The second stage of elicitation breaks these ties: it is now strictly optimal for participants who are genuinely indifferent between two or more options to say so.

The effect of the multi-stage elicitation procedure on one category of respondent is

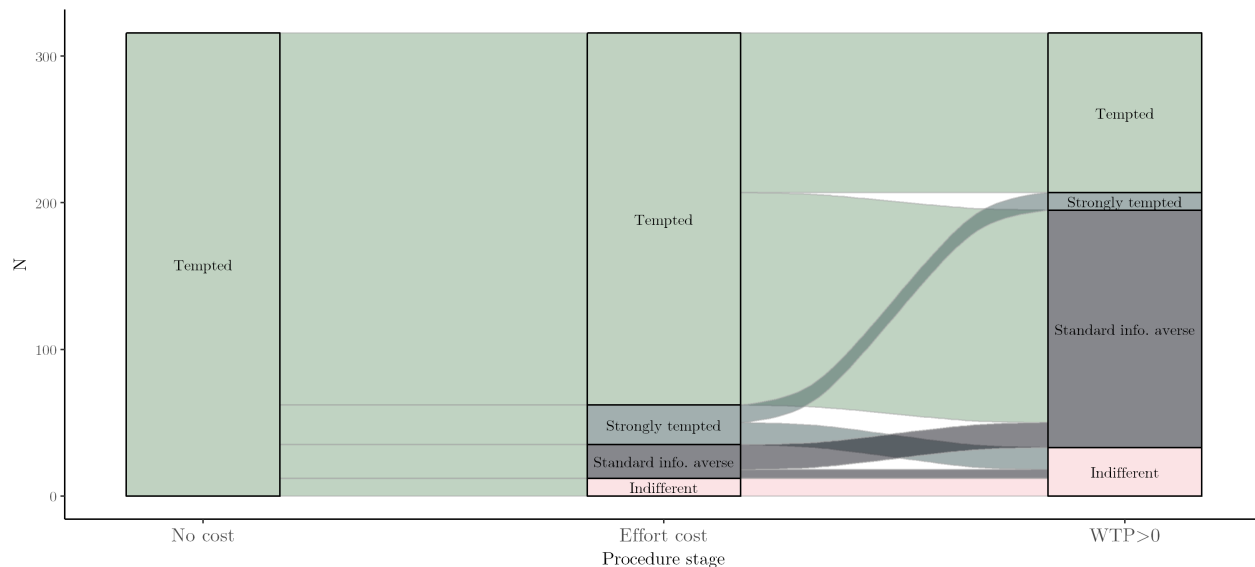
---

<sup>7</sup>As we mentioned, we report all our main results across the whole sample, including both participants who took the bonus and those who did not. The results are qualitatively similar if just considering the subset of participants who took the bonus (Table A3).

<sup>8</sup>In all, just over two-thirds of the sample strictly prefers to avoid the information offered than see it (Table A1).

<sup>9</sup>Type categories are condensed in Table A4.

Figure 2: Preference transitions over the elicitation procedure: types classified as ‘tempted’ in the no cost stage.



*Note:* ‘No cost’ refers to the stage where participants simply enter their preference rankings into a grid. ‘Effort cost’ is our main measure of preferences, and is taken after participants must complete a short task to confirm each adjacent strict preference in their ranking.  $WTP > 0$  captures those who are willing to pay money for their first adjacent strict preference.

illustrated in Figure 2. We see that a large majority (80%) of those with the preference ordering  $\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$  before the effort cost stage are willing to pay effort costs to maintain both strict preferences. A roughly equal number of those unwilling to pay the effort cost are subsequently reclassified as strongly tempted ( $\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$ ) or standard information averse ( $\{0\} \sim_1 \{0, 1\} \succ_1 \{1\}$ ) types. In the first case, participants state that they are willing to complete a short task to confirm their preference for ‘not generating the envelope at all’ over ‘deciding whether to open the envelope next time’, but unwilling to do the same to confirm their preference for ‘deciding whether to open the envelope next time’ over ‘automatically opening the envelope next time’. In contrast, those reclassified as standard information averse types are unwilling in the first case but willing in the second case. A smaller share are unwilling to pay either effort cost and are thus classified as indifferent.

There may be some concern that the initial procedure, involving asking participants to select ranks 1, 2, and 3 for each alternative, naturally encourages participants to submit strict preferences. One reason for this could be that the joint consideration of three alternatives is cognitively demanding. Our main measure addresses this issue by prompting participants to reconsider the adjacent *pairs* of options for which they initially submitted strict preferences. Any participant with a strict preference for commitment over flexibility in our main measure must have explicitly confirmed they had a clear preference for ‘not generating the envelope’ over ‘deciding whether to open the envelope next time’ (Figure A6), and completed a short



task to do so.

Since effort costs are mandatory for strict preferences in our main measure, it is possible that we measure some genuine strict preferences as indifferences where participants simply do not value their preference at the margin above the effort cost we ask them to pay. Overall, 29% of participants with initial strict preferences decline to pay at least one of the unit of the effort cost in proceeding to the next stage. In this sense, this measure of preferences — our preregistered primary outcome — can be seen as conservative on balance.

Even more conservative is the third and final measure, which requires participants to have a strictly positive monetary *WTP*, having already paid effort costs, to keep their first and second choices in place. Some other studies of commitment devices, such as Augenblick et al. (2015), find that demand drops to nearly 0 when participants are asked if they are willing to pay money. In our case, about 43% of tempted types are also willing to pay strictly positive sums of money in order for  $\{0\}$  and  $\{0, 1\}$  not to be swapped in their preference rankings, even though they have already completed two effort tasks confirming their strict preferences (Table A5). This leaves 16% of individuals as willing to pay both effort costs and small sums of money in order for information not to be generated and offered to them. Following this stage, the remaining 57% of tempted types would be classified as standard information averse. However, given that all of these individuals already paid an effort cost to justify their strict preference by this stage, this classification is very conservative and we do not emphasise it. Instead, it can be seen as a sort of lower bound on the prevalence of strict preferences for commitment, particularly for readers who are not convinced that effort costs are sufficient to extract genuinely strict preferences for commitment over choice.

The average *WTP* among tempted types is \$0.09, reflecting the high share with  $WTP=0$  (Table A5).<sup>10</sup> Among those with a positive *WTP*, the mean is \$0.20. Strongly tempted types express similar valuations, with 45% having  $WTP > 0$  and an average *WTP* of \$0.26 among those with  $WTP > 0$ . There is some evidence of bunching at \$0.50 across both types, suggesting that we underestimate average *WTP*: 17% of tempted and strongly tempted types with a positive *WTP* submit the maximum *WTP* of \$0.50. Otherwise, *WTP* is somewhat evenly distributed across the range of permitted values (Figure 3).<sup>11</sup>

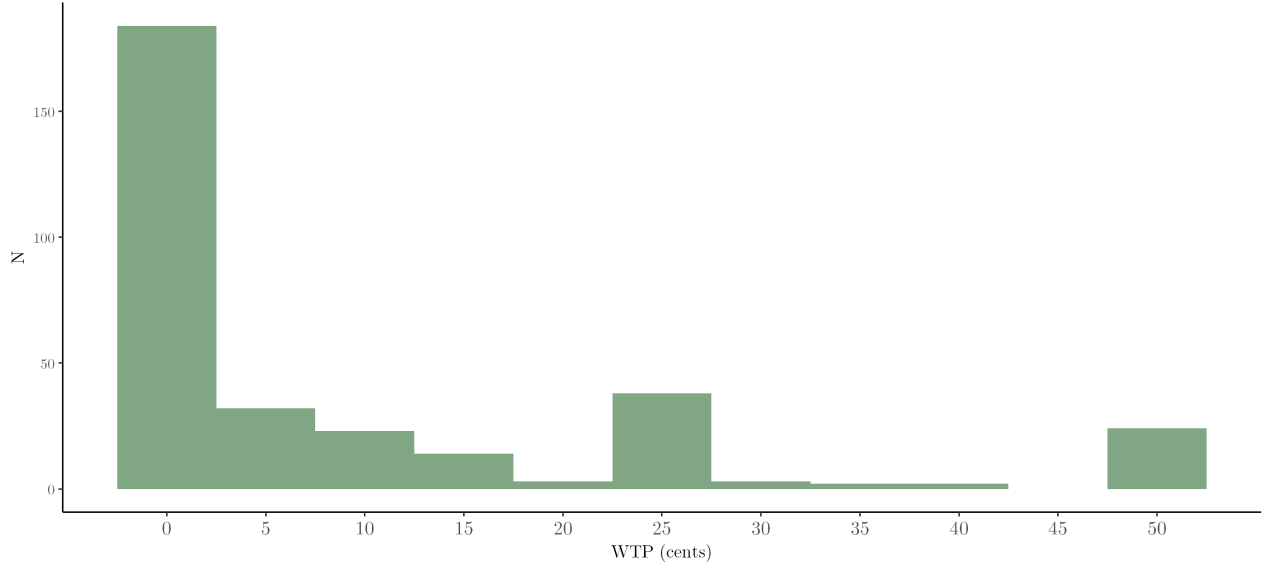
Our data also include a set of participant characteristics which we can use to examine

---

<sup>10</sup> *WTP* is bounded above at \$0.50, and participants could only submit multiples of 5 cents (\$0, \$0.05, \$0.10, ..., \$0.50).

<sup>11</sup> For other types,  $WTP > 0$  has a different interpretation: since  $\{1\}$  or  $\{0\}$  appears as (joint) second-ranked for many of these individuals, they are asked for their *WTP* to avoid an order swap between seeing and not seeing the information. Taking, for instance, individuals with  $\{0\} \sim_1 \{0, 1\} \succ_1 \{1\}$ , the question posed was how much money they were willing to forgo to avoid a swap between  $\{0\}$  and an alternative randomly selected from  $\{0, 1\}$  and  $\{1\}$ . The swap entails a possibility of automatically opening the envelope in session 2 (and being charged for doing so), rather than simply being offered the envelope.

Figure 3: *WTP* histogram: tempted and strongly tempted types



*Note:* Willingness to pay for the menu preference  $\{0\} \succ_1 \{0, 1\}$ , elicited via the BDM procedure. If the participant lost the auction, their new menu preference ranking would begin with  $\{0, 1\} \succ_1 \{0\}$ .

determinants of preference heterogeneity. Participants' age, ethnicity and sex were automatically provided by the panel, whereas we manually elicited qualitative measures of risk preferences, patience, self-control and altruism at the end of session 2.<sup>12</sup> Since the range of preference rankings is broad, we report two regressions focusing on cases of particular interest. The first predicts the event of an implied binary preference to avoid information,  $\{0\} \succ_1 \{1\}$ . In the first case, the likelihood of submitting a strict preference to avoid information is lower among some older age categories and participants of Asian ethnic origin. Other variables are not strong predictors of preferences to avoid information, notably including the price of information and self-reported scores relating to altruism, patience, and self-control. The exception is that participants with higher subjective risk tolerance are less likely to submit preferences to avoid information. This is likely a reflection of the fact that the decision to seek information is inherently risky, given that it can contain either good or bad news.

The second regression is for the probability of  $\{0\} \succ_1 \{0, 1\}$  conditional on  $\{0\} \succ_1 \{1\}$ . Very few of the characteristics in our data have any predictive power. The exception is a weakly positive correlation of commitment demand with age.

## 4.2 Information choices from session 2

**Result 2:** *3% of individuals tempted by undesired information pay to access it in session 2.*

<sup>12</sup>Since these measures were collected after bonus and information choices, they could be endogenous to both. As such, this analysis is presented descriptively.

In session 2, taking place a day after session 1, participants' preferences are randomly implemented according to the scheme previously described. With 50% probability, menu preferences were ignored and  $\{0, 1\}$  is implemented. With 40% probability, the top menu from  $\succeq_1$  was implemented. With 10% probability, the second menu from  $\succeq_1$  is implemented. Of the segment facing  $\{0, 1\}$  exogenously, 16% elect to pay to open the envelope overall.

Table 3 splits session 2 information choices by implied session 1 binary rankings of  $\{0\}$  and  $\{1\}$ . Information access is markedly more common for participants who submitted rankings with  $\{1\} \succ_1 \{0\}$  than those with  $\{1\} \sim_1 \{0\}$ . Matching other experimental studies of dynamic consistency for other choice contexts, succumbing to temptation is relatively uncommon. However, it is notable that some individuals do still break their preference in this fashion, given that they must now pay a small sum of money to access the information having borne at least an effort cost not to be offered it only a day earlier. For those individuals who submitted  $\{1\} \succ_1 \{0\}$ , 65.6% follow through on their session 1 preference when asked to choose again. However, it should be noted that a majority (56%) of participants in this category weakly preferred flexibility to committing to seeing the information, meaning their session 1 preference for  $\{1\}$  over  $\{0\}$  is likely to have been relatively uncertain.

Table 3: Session 2 information choice by session 1 information preference

Session 1 information preference	% $1 \succ_2 0$	...of $N$
$\{0\} \succ_1 \{1\}$	5.4% (1.7)	184
$\{0\} \sim_1 \{1\}$	21.6% (5.9)	51
$\{1\} \succ_1 \{0\}$	65.6% (8.6)	32
Total	15.5% (2.2)	265

*Note:* Standard errors in parentheses, in percentage points. Only includes participants who were randomly allocated to receive  $\{0, 1\}$  regardless of their menu preferences.

Table 4 provides a more granular separation of session 2 choices by session 1 preferences. We first focus on the two main types of interest: tempted and strongly tempted types, both of whom view information as tempting and strictly prefer not to be offered it as a result. Preference reversals are relatively infrequent across both types, at rates of 2% for tempted types and 7% for strongly tempted types. As we will see later, individuals who are more likely to succumb to temptation are less willing to pay an effort cost for the strict preference  $\{0, 1\} \succ_1 \{1\}$ , meaning they are classified as strongly tempted types in the effort cost stage of our elicitation procedure.

Among types with  $\{0\} \succ_1 \{1\}$ , the highest probability of accessing information is among those with the rankings  $\{0\} \succ_1 \{1\} \succ_1 \{0, 1\}$  and  $\{0, 1\} \succ_1 \{0\} \succ_1 \{1\}$ . This first type could be classified as 'flexibility averse', although the group exogenously offered  $\{0, 1\}$  is very

Table 4: Session 2 information choice by session 1 menu preference

Session 1 menu preference	Type	% $1 \succ_2 0$	...of $N$
$\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$	Tempted	1.9% (1.3)	105
$\{0\} \sim_1 \{0, 1\} \sim_1 \{1\}$	Indifferent	23.5% (7.3)	34
$\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$	Dynamic inconsistency	6.9% (4.7)	29
$\{0\} \sim_1 \{0, 1\} \succ_1 \{1\}$	Standard info. averse	8% (5.4)	25
$\{0, 1\} \succ_1 \{0\} \succ_1 \{1\}$	Flex	15.8% (8.4)	19
$\{0, 1\} \succ_1 \{0\} \sim_1 \{1\}$	Flex	7.7% (7.4)	13
$\{0, 1\} \succ_1 \{1\} \succ_1 \{0\}$	Flex	80% (12.6)	10
$\{0, 1\} \sim_1 \{1\} \succ_1 \{0\}$	Other	42.9% (18.7)	7
$\{1\} \succ_1 \{0, 1\} \succ_1 \{0\}$	Tempted (info. loving)	83.3% (15.2)	6
$\{0\} \succ_1 \{1\} \succ_1 \{0, 1\}$	Flexibility averse	16.7% (15.2)	6
$\{1\} \succ_1 \{0\} \sim_1 \{0, 1\}$	Other	60% (21.9)	5
$\{0\} \sim_1 \{1\} \succ_1 \{0, 1\}$	Commitment loving	66.7% (27.2)	3
$\{1\} \succ_1 \{0\} \succ_1 \{0, 1\}$	Flexibility averse	33.3% (27.2)	3
Total			265

*Note:* Standard errors in parentheses, in percentage points. Only includes participants who were randomly allocated to receive  $\{0, 1\}$  regardless of their menu preferences.

small, comprising 6 individuals. The second type is a little larger, with 19 individuals, and encompasses those with a strict preference for flexibility. The preferences of these individuals are easier to interpret: although they prefer  $\{0\}$  to  $\{1\}$  under  $\succeq_1$ , they strictly prefer flexibility to either form of commitment. In that sense, either  $1 \succ_2 0$  or  $0 \succ_2 1$  is compatible with  $\succeq_1$  for these individuals.

### 4.3 Self-control costs

**Result 3:** *Realised self-control costs are strongly related to preferences for commitment: strongly tempted types spend longer deciding whether to access information than tempted types; and individuals who are initially willing to pay more for commitment eventually have longer deliberation periods. Longer deliberation periods increase the probability of accessing information.*

So far, we have established that a substantial proportion of individuals in our experiment believe they will suffer from the mere presence of information in their choice set. 49% of our sample are willing to pay an effort cost to confirm a strict preference for restricting their choice set, and 43% of those individuals are also willing to pay small sums of money for the same purpose. As in Toussaert (2018), dynamically inconsistent choice or random

self-indulgence (Dekel and Lipman, 2012) are unlikely to explain all of the demand for commitment devices; individuals rarely succumb to temptation in session 2. Our model suggests that a significant role must therefore be played by  $m_2$ , the self-control cost from having to turn down information when it is offered in session 2.

Online delivery of the experiment permitted us to collect a naturally occurring measure of self-control costs in session 2: time spent on the page where the choice is made to open the envelope. Participants spent on average around 16 seconds on the page when  $\{0, 1\}$  was exogenously implemented. However, more instructive is how self-control costs differ across session 1 menu preferences, which were elicited a day earlier, as well as across individuals with different willingness to pay effort costs and money for their preferences for commitment.

**Types and self-control costs.** Our model of temptation and self-control predicts a relationship between menu preferences and expected self-control costs. From Proposition 1, within the set of individuals who find information tempting, ‘strongly tempted’ types ( $\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$ ) should experience higher costs of self-control than ‘tempted’ types ( $\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$ ) when facing the menu  $\{0, 1\}$ .

Mean deliberation time is relatively short for tempted types, at about 13 seconds (Table 5).<sup>13</sup> On the other hand, strongly tempted types spent around 5 seconds longer on average deciding whether or not to open the envelope. Deliberation time was relatively short for standard information averse types, who should have both low self-control costs and a weak *ex ante* preference for avoiding information.<sup>14</sup>

Thus, realised self-control costs closely match the predictions of Proposition 1. Recall that the difference between self-control and strongly tempted types is that the latter anticipates either relatively high costs of self-control or a high probability of facing overwhelming temptation. The data suggest that both concerns reflect some sophistication a day ahead of the decision: strongly tempted types spend significantly longer deciding whether to open their envelopes than tempted types, and are also more likely to actually open them.

An open question stemming from the above is whether a longer period of deliberation predicts failure of self-control. We estimate a logistic regression for the event that the envelope is opened in session 2 under exogenously implemented  $\{0, 1\}$ , with quadratic deliberation time as the regressor of interest, for those whose session 1 preferences imply they find information tempting (Table A8). While the coefficients on decision time are not individually significant, they are jointly significant at the 5% level. The log-odds of failed self-control increase

---

<sup>13</sup>deliberation times for all types are reported in the Appendix, in Table A7.

<sup>14</sup>The longest deliberation times were among types who strictly preferred flexibility, suggesting the interpretation that these types were particularly uncertain of their preferences for the information.

Table 5: Measures of deliberation under exogenous  $\{0, 1\}$  in session 2 by session 1 preference ranking

Menu preference	Type	Time (seconds)	$N$
$\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$	Tempted	13.1 (0.5)	105
$\{0\} \sim_1 \{0, 1\} \sim_1 \{1\}$	Indifferent	17.5 (1.9)	33
$\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$	Strongly tempted	18 (1.9)	28
$\{0\} \sim_1 \{0, 1\} \succ_1 \{1\}$	Standard info. averse	15.6 (2.3)	25
$\{0, 1\} \succ_1 \{0\} \succ_1 \{1\}$	Flex	21.5 (2.9)	19
$\{0, 1\} \succ_1 \{0\} \sim_1 \{1\}$	Flex	25.9 (3.2)	13
$\{0, 1\} \succ_1 \{1\} \succ_1 \{0\}$	Flex	16.6 (1.7)	10

*Note:* Only the most common types included, for brevity. Full table in Appendix.

substantially in deliberation time, before levelling off and decreasing slightly at very long deliberation periods.<sup>15</sup>

**Willingness to pay and self-control costs.** Our model also generates granular predictions on how the *intensity* of individuals' preferences for commitment are determined by expectations of self-control costs and the probability of failed self-control (Corollary 1). Recall that the strength of the preference  $\{0\} \succ_1 \{0, 1\}$  is determined by  $(1 - \tilde{p})m_1 + \tilde{p}m_2 - c$ . The strength of the preference  $\{0, 1\} \succ_1 \{1\}$  is determined by  $\tilde{p}(m_1 - m_2) - 4c$ .

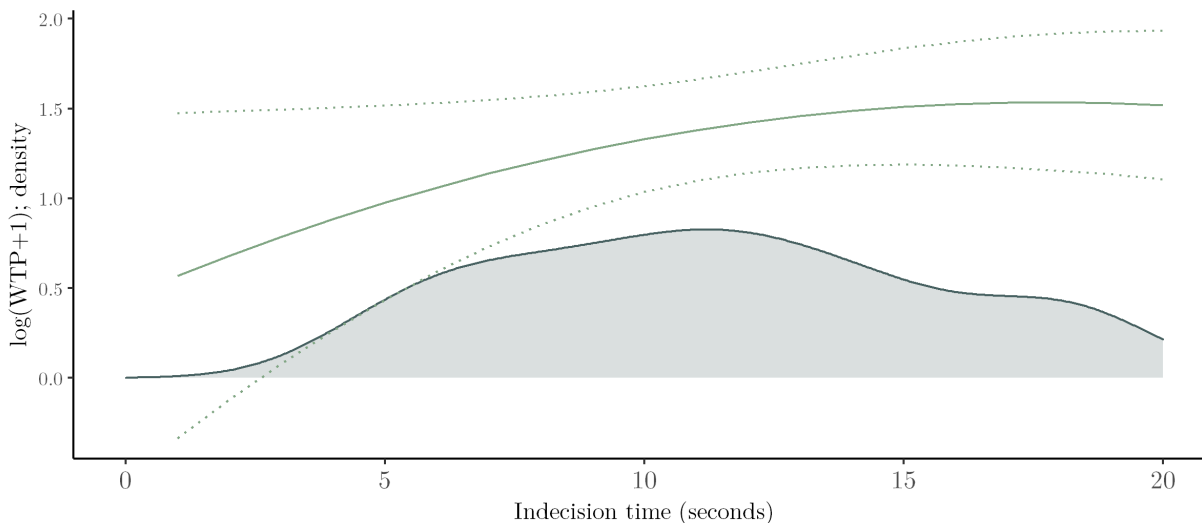
For this purpose, our multi-step elicitation procedure provides two separate measures of how much individuals value their menu preferences. In the main stage, we obtain a binary measure of whether an individual is willing to bear an effort cost to maintain their strict preferences. After that, we collect a continuous measure of individuals' *WTP* to avoid a switch between their first and second strictly ranked menus.

We can test these predictions using approximate measures of the parameters. Our approximate measure of  $\tilde{p}$  is simply a dummy variable denoting whether the individual opened the envelope ( $a = 1$ ) when  $\{0, 1\}$  was exogenously implemented in session 2. Our measure of  $m_1$  is each individual's stated interest, at the end of session 1, in accessing the information. Finally, we continue to measure  $m_2$  using each individual's quadratic deliberation time in session 2 when  $\{0, 1\}$  is implemented. We include these variables in regressions predicting whether individuals are willing to pay effort costs to maintain  $\{0\} \succ_1 \{0, 1\}$  and  $\{0, 1\} \succ_1 \{1\}$ , as well as monetary *WTP* in the former case.

The results are reported in Table A9. We first consider willingness to pay effort costs and money for  $\{0\} \succ_1 \{0, 1\}$ , which our model predicts to increase in  $(1 - \tilde{p})m_1 + \tilde{p}m_2$ . Column

<sup>15</sup>The probability of accessing information reaches its peak at just over 15 seconds of deliberation and decreases after that point only slightly.

Figure 4: Period 2 deliberation time vs predicted period 1  $WTP$  to maintain  $\{0\} \succ_1 \{0, 1\}$  for tempted types



Note: Scaled density of decision times plotted underneath predicted  $WTP$ . 95% prediction interval plotted around predicted  $WTP$ .

(1) contains the results for effort costs, while (2) examines monetary  $WTP$ . Our measure of the marginal commitment payoff to avoiding information,  $m_1$ , does not significantly impact the willingness to pay either effort costs or money. More granular is our continuous measure of self-control costs in period 2, given by deliberation time under exogenously implemented  $\{0, 1\}$ . For this variable, we obtain a positive effect on the willingness to pay both effort costs and money (columns (1) and (2)), although the coefficient on the linear effect is statistically significant only in the latter case.

Most striking is the result from column (2): within the set of those tempted by information, those willing to pay larger amounts of money to maintain  $\{0\} \succ_1 \{0, 1\}$  generally appear to find it harder to resist information a day later. This is precisely as predicted by our model of costly self-control: larger anticipated self-control costs  $m_2$  make the preference  $\{0\} \succ_1 \{0, 1\}$  stronger. Since the effect of self-control costs on  $WTP$  is non-linear, it can be more easily observed visually, as in Figure 4. Those with higher self-control costs in period 2, are willing to pay more money to maintain their preference for commitment over flexibility,  $\{0\} \succ_1 \{0, 1\}$ . As we approach relatively high deliberation time periods (recall from Table 5 that the average for this type is 13.1 seconds), the effect of deliberation time on  $WTP$  begins to level off and decrease gently. The effect of deliberation time is economically significant. Roughly, a 10-second increase in deliberation time predicts a 70% increase in willingness to pay for  $\{0\} \succ_1 \{0, 1\}$ .

We finally consider column (3), predicting willingness to pay effort costs to maintain the

Table 6: Regressions for preference intensity

	$\{0\} \succ_1 \{0, 1\}$		$\{0, 1\} \succ_1 \{1\}$
	Paid effort cost (1)	$\log(WTP + 1)$ (2)	Paid effort cost (3)
Not at all interested ( $m_1$ )	0.84 (0.53)	-0.28 (0.30)	-0.06 (0.45)
Time ( $m_2$ )	0.12 (0.09)	0.12 (0.06)	-0.04 (0.07)
Time <sup>2</sup> ( $m_2$ )	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
Opened envelope ( $1 - \tilde{p}$ )	-0.15 (1.23)	0.54 (0.77)	-1.57 (0.97)
Constant	0.60 (0.91)	0.64 (0.56)	1.98 (0.80)
Observations	149	133	149

*Note:* Coefficients are log-odds, other than for column (2) which is OLS. Sample restricted to those randomised to face  $\{0, 1\}$  exogenously in session 2. Columns (1) and (3) capture all those tempted by information in the raw elicitation stage, while (2) includes only those classified as tempted or strongly tempted in the main stage.

preference  $\{0, 1\} \succ_1 \{1\}$ .<sup>16</sup> The corresponding condition from our model of costly self-control is  $\tilde{p}(m_1 + m_2) - 4c > 0$ . Here, neither  $m_1$  nor  $m_2$  significantly determine the willingness to pay effort costs. However, our crude measure of  $1 - \tilde{p}$  has a significant, negative effect: individuals who do eventually succumb to temptation in period 2 are significantly less willing to pay effort costs for the strict preference  $\{0, 1\} \succ_1 \{1\}$ . Our model dictates that this is true only if  $m_1 > m_2$ ; otherwise, the effect of  $\tilde{p}$  on the intensity of the preference for  $\{0, 1\}$  over  $\{1\}$  is 0. Thus, in this case, lower values of  $\tilde{p}$  make flexibility less attractive relative to automatically accessing information. Intuitively, if successful self-control is unlikely, it is not worth bearing effort costs to have the ability to avoid information.

#### 4.4 Comparison of treatment conditions

**Result 4:** *Unavoidable partial information erodes preferences to commit to avoiding unwanted information.*

Since the total content and average price of information is identical across treatment conditions, our prior analysis considered menu preferences and information choices aggregated across the whole sample. In reality, individuals who take the bonus are randomly allocated to one of two treatment conditions: *PartAvoid* and *FullAvoid*. To recall, there is only a consequence of taking the bonus if the individual is a *Donor*. Participants in both treatments are able to confirm if they were a *Donor* during session 2. What differs is how much information

<sup>16</sup>Since this was the second adjacent strict preference for tempted types, we did not elicit monetary *WTP* for it.



participants can choose to avoid in session 2, and correspondingly what they can commit to avoiding in session 1.

In *PartAvoid*, participants cannot avoid receiving an initial signal of the consequences of their decision to take the bonus. The signal  $s$  takes values  $s_1$  or  $s_2$ , where  $P(\text{Donor}|s = s_1) = 0$  and  $P(\text{Donor}|s = s_2) = 0.2$ . If  $s = s_2$ , an envelope letting the participant learn the value of *Donor* is generated. Recalling Figure 1, the participant then chooses to open or not open the envelope, denoted by the action  $a_c$ . Session 1 menu preferences in *PartAvoid* thus correspond to the set of menus  $\mathcal{M}_c \in \{\{0\}, \{0, 1\}, \{1\}\}$ , relating to  $a_c$ .<sup>17</sup> In *FullAvoid*, the signal  $s$  can be avoided through an action  $a_s$ , so participants always have an opportunity to obtain strictly less information than those in *PartAvoid* by setting  $a_s = 0$ . Session 1 menu preferences in *FullAvoid* thus correspond to the set of menus  $\mathcal{M}_s \in \{\{0\}, \{0, 1\}, \{1\}\}$ , relating to  $a_s$ . If  $a_s = 1$  and  $s = s_2$ ,  $a_c$  must be set in session 2, but no choice set restrictions are offered for  $a_c$  so that participants in both treatments only need to submit one set of menu preferences in session 1. The only difference across the two treatments in session 1 consists of a small variation in the text explaining the information offer sequence and the contents of the envelope (Figures A3 and A4). The text describing the choice alternatives and the subsequent stages of the elicitation procedure only make reference to the envelope, so they are identical across the two treatments.

Differences in menu preferences across the treatment conditions measure whether it is easier or harder to commit to avoiding information when more of it must be avoided. If commitment demand is higher in *FullAvoid* than *PartAvoid*, we can conclude that avoiding more information makes commitment less attractive; in other words, initial information is tempting because of the higher chance it provides good news. If commitment demand is higher in *FullAvoid*, we can conclude that committing to avoid information is harder when partial information cannot be avoided.

Table 7 splits session 1 menu preferences by treatment group. There is no significant difference across treatment groups in the prevalence of tempted types. However, the share of strongly tempted types, with menu preference  $\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$ , is marginally greater in *FullAvoid* than *PartAvoid*. This type strictly prefers commitment but is indifferent between being offered the information and automatically seeing it. Relative to tempted types, of which there is no difference in prevalence across treatment arms, strongly tempted types rank the menus  $\{0, 1\}$  and  $\{1\}$  equally. As noted in Proposition 1, one possibility is that strongly tempted types have a weaker *ex ante* preference for avoiding information. Since

---

<sup>17</sup>Note that all the menus in *PartAvoid* are only offered if  $s = s_2$ , in order to avoid eliciting multiple conditional menu preferences. While this could reduce participants' willingness to pay for their strict preferences, we find no evidence of this phenomenon: there is little difference in participants' willingness to pay effort costs to confirm their strict preferences.

Table 7: Menu preferences: treatment comparison

Menu preference	Type	N		Share		Diff.	p-value
		<i>PartAvoid</i>	<i>FullAvoid</i>	<i>PartAvoid</i>	<i>FullAvoid</i>		
$\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$	<b>Tempted</b>	118	113	<b>37.7%</b> (2.7)	<b>39.1%</b> (2.9)	-1.4	0.79
$\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$	<b>Strongly tempted</b>	26	38	<b>8.3%</b> (1.6)	<b>13.1%</b> (2)	-4.84	0.07
$\{0\} \sim_1 \{0, 1\} \sim_1 \{1\}$	Indifferent	43	30	13.7% (1.9)	10.4% (1.8)	3.36	0.26
$\{0\} \sim_1 \{0, 1\} \succ_1 \{1\}$	<b>Standard info. averse</b>	21	27	<b>6.7%</b> (1.4)	<b>9.3%</b> (1.7)	-2.63	0.3
$\{0, 1\} \succ_1 \{0\} \succ_1 \{1\}$	Flex	23	25	7.3% (1.5)	8.7% (1.7)	-1.3	0.66
$\{0, 1\} \succ_1 \{1\} \succ_1 \{0\}$	Flex	13	16	4.2% (1.1)	5.5% (1.3)	-1.38	0.55
$\{1\} \succ_1 \{0, 1\} \succ_1 \{0\}$	Strongly tempted (info. loving)	15	11	4.8% (1.2)	3.8% (1.1)	0.99	0.69
$\{0, 1\} \succ_1 \{0\} \sim_1 \{1\}$	<b>Flex</b>	23	8	<b>7.3%</b> (1.5)	<b>2.8%</b> (1)	4.58	0.02
$\{0, 1\} \sim_1 \{1\} \succ_1 \{0\}$	Other	9	6	2.9% (0.9)	2.1% (0.8)	0.8	0.71
$\{1\} \succ_1 \{0\} \sim_1 \{0, 1\}$	Other	8	5	2.6% (0.9)	1.7% (0.8)	0.83	0.68
$\{0\} \succ_1 \{1\} \succ_1 \{0, 1\}$	Flexibility averse	7	5	2.2% (0.8)	1.7% (0.8)	0.51	0.88
$\{0\} \sim_1 \{1\} \succ_1 \{0, 1\}$	Commitment loving	5	3	1.6% (0.7)	1% (0.6)	0.56	0.81
$\{1\} \succ_1 \{0\} \succ_1 \{0, 1\}$	Flexibility averse	2	2	0.6% (0.5)	0.7% (0.5)	-0.05	1
Total		313	289	100%	100%		

*Note:* The 9% of participants who did not take the bonus are excluded from this table, since they were all allocated to the same condition (in which all of the information was accessed at the same time). Difference in percentage points.

in *PartAvoid*, the proposition of avoiding information ( $\{0\}$ ) involves having already learned there was a 20% chance of being a *Donor*, it appears to become less attractive to do so. This also explains the finding that the increased prevalence of strongly tempted types in *FullAvoid* appears to be at the expense of flexibility types with  $\{0, 1\} \succ_1 \{0\} \sim_1 \{1\}$ , who have the same preference except with  $\{0\}$  and  $\{0, 1\}$  switched. This type is consistent with preference uncertainty: they are indifferent between  $\{0\}$  and  $\{1\}$  under  $\succeq_1$ , and so would like to wait to make their decision. These conclusions are also borne out when condensing menu preferences into binary information preferences, as in Table A10.

Table 8 examines differences in session 2 behaviour across treatment groups. This analysis is underpowered: only around half of participants are allocated to the condition in which  $\{0, 1\}$  is exogenously implemented, so splitting by treatment group results in a very small sample. A higher proportion of individuals in *FullAvoid* access the information than in *PartAvoid*, but the corresponding standard errors are about 3 percentage points.

To sum up, information choices in *PartAvoid* are made only after unfavourable partial information has already been received. It is striking that individuals are nonetheless *ex ante* less able to resist acquiring further information. This suggests the provision of partial information could be an effective strategy in tempting access to further information: it is harder for individuals to plan to avoid even potentially upsetting information when they expect to be given an initial indication of its content.

Table 8: Session 2 information choice from exogenous  $\{0, 1\}$  by treatment group

<i>N</i>		Share			
<i>PartAvoid</i>	<i>FullAvoid</i>	<i>PartAvoid</i>	<i>FullAvoid</i>	Diff.	<i>p</i> -value
108	157	13% (3.2)	17.2% (3)	-4.2	0.44

## 4.5 Qualitative measures of information preferences and self-control costs

We collected a set of qualitative measures in both sessions, which we now use to further interrogate the motives for choice set restrictions.

We begin by more closely examining our curiosity measure ‘How interested are you to know if the charity donation was reduced by \$15 as a result of your choice?’, which we previously introduced as a measure of  $m_1$  in our analysis of willingness to pay. In session 1, self-control, weak self-control and standard information averse types are those with the lowest measured curiosity, which mirrors the idea from our model that the commitment evaluation of avoiding information  $m_1$  should be highest among these types. Curiosity in session 1 is much higher among flexibility types and those indifferent between all menus.

Since we asked this question in both sessions, within-individual comparisons reveal how  $m_2$ , the temptation utility evaluation of avoiding information, differs from  $m_1$ . Noting that we only include participants who successfully resisted information in period 2, tempted types experience a significant increase in the subjective attractiveness of information from session 1 to session 2. Thus, tempted types seem to be justified in their revealed preference for commitment, even though succumbing to temptation is relatively rare. The increase in interest in information is not statistically significant for strongly tempted types, although the test is poorly powered. Standard information averse types do also experience a significant increase in the attractiveness of information. Since these types were unwilling to pay effort costs to commit to avoiding information, the increase in the attractiveness of information could suggest some naivete about self-control costs at the time of submitting menu preferences. Finally, flexibility types generally did not experience increases in the attractiveness of information. Indeed, those with the ranking  $\{0, 1\} \succ_1 \{0\} \sim_1 \{1\}$  experience a sizeable *decrease* in the attractiveness of information conditional on deciding not to access it.

We can also draw on participants’ own rationales for their session 1 menu preferences. In all, 69% of participants chose to answer a free-text question in session 1 rationalising their preference rankings. Most respondents chose to focus on their binary information preference  $\{0\} \succ_1 \{1\}$ , reflecting the strength of sentiments about the \$4 choice and the desire to avoid information about its consequences. Overwhelmingly, responses within this set appear to

indicate a desire to engage in wishful thinking about the consequences of their action or simply avoid thinking about it at all.

Perhaps more interesting is the set of responses for tempted types which explicate reasons for submitting  $\{0\} \succ_1 \{0, 1\}$ . Although it is often difficult to distinguish clearly between these two interpretations, about half allude to self-control costs (e.g. ‘I would like to remove the option because it will help me forget feeling bad about my choice.’; ‘I’d rather not know and not have to make the decision.’) and half seem to perceive a risk of failed self-control (e.g. ‘I would rather not be given the option so I don’t feel tempted to open it and find out.’; ‘Because that way I don’t even try to open it at all’). This is consistent with our previous findings: at least some individuals correctly believe they have a chance of later accessing information in spite of their preferences, in line with a model of random self-indulgence (Dekel and Lipman, 2012), but many others seem to be sure they will not open the envelope even if it is offered next time, suggesting costly self-control must be the key motive. In the latter category, several participants mention that being offered the envelope would prolong the guilt they feel about their decision to take the bonus. This suggests avoiding information may be psychologically costly, perhaps because it results in cognitive dissonance. We harness this idea in a simple model of information preferences, set out in Appendix A3.

## 5 Discussion

Our results imply that information consumption can be driven by instant gratification rather than classical utility maximisation. They also suggest that partial information may tempt access to further information. As such, they have implications spanning both theoretical and empirical research on intrinsic preferences for information, as well as providing a starting point for policy discussions on markets for information.

### 5.1 Implications for theory

The Gul and Pesendorfer (2001) model of temptation and self-control can be interpreted as reflecting dynamic or intra-self inconsistency in preferences (Bénabou and Pycia, 2002). Indeed, the Gul and Pesendorfer (2001)-style representation we use in Section 2 relies on the use of two local utility functions  $u_1$  and  $u_2$  with directly conflicting evaluations of information.  $u_2$  can be interpreted as the instantaneous preference for information, whereas  $u_1$  is the *ex ante* preference.

Our data lead us to the conclusion that intrinsic preferences for information may be dynamically inconsistent for many individuals. A small share of individuals directly demonstrate

dynamic inconsistency by paying to access information after having strictly preferred to not be offered it. However, preferences also appear dynamically inconsistent at the *intensive* margin: using our behavioural measure of self-control costs, we find evidence that for tempted types, session 2 preferences for information are negatively correlated with session 1 (*ex ante*) preferences for information. Qualitative measures also indicate that participants we classify as tempted by information are more interested in seeing it in session 2, when it is available, than in session 1.

All existing theories of intrinsic preferences for information would require some extension to accommodate dynamic inconsistency. We briefly consider what this might entail for some major categories of models below.

**Anticipatory utility.** Models in the vein of Caplin and Leahy (2001, 2004), Brunnermeier and Parker (2005), and Ely et al. (2015) generate information avoidance from a desire to avoid unpleasant feelings about past or future outcomes. A particular focus of Caplin and Leahy (2001), and one that has drawn significant interest from subsequent theoretical and empirical research, is on the role of anxiety in driving information demand.

Within the lens of these models, our results would suggest that accessing unwanted information supplies affective instant gratification at the cost of long-term welfare. While one can supply many rationales, one example could be that ‘not knowing’ creates unpleasant feelings which reach a fever pitch at the time information can be accessed, but subsequently dissipate. Under present bias, individuals may demand information in order to assuage unpleasant feelings at the cost of learning something they would rather not know. Commitment devices may permit curiosity to be mitigated because attention can be diverted away from the subject of information (Golman et al., 2021; Falk and Zimmermann, 2023).

**Reference-dependent utility.** Kőszegi and Rabin (2009) consider a model in which individuals’ utility depends on their loss-averse beliefs about future outcomes. In this setting, individuals avoid information because they are more averse to bad news than good news. Thus, time inconsistency in the demand for information would require time inconsistency in preferences. However, the authors suggest that many of their key results could be affected by relaxing time consistency.

**Preferences for the timing of uncertainty resolution.** Another set of models provides axiomatic foundations for information demand by allowing individuals to have preferences over the timing of the resolution of uncertainty (Kreps and Porteus, 1978; Epstein and Zin, 1989; Grant et al., 1998; Epstein, 2008). In these models, attitudes towards information

result from fundamental features of preferences, such as the concavity or convexity of local utility. Thus, temptation over information would imply time or intra-self inconsistency in these fundamental features of preferences. This may be most natural in models where information preferences relate to intertemporal substitution parameters (Epstein and Zin, 1989).

Aside from demonstrating dynamic consistency in intrinsic preferences for information, our results confirm the conjecture of many authors that avoiding information could be inherently costly to welfare. This has major implications for the theoretical analysis of information avoidance, including in settings where information has instrumental value. For example, the core models in Bénabou and Tirole (2002) and Brunnermeier and Parker (2005) treat information avoidance as intrinsically costless: the only consequence of avoiding information is on decision making. Our results suggest that a different approach may be warranted: the self-control costs incurred in resisting information are economically significant.

## 5.2 Implications for empirical research

Broadly, our results suggest that more care may be required in empirical measurements of information preferences. The fact that some preference reversals were observed in our experiment is fairly striking given individuals had to pay to access information and made their choices in a highly controlled environment. It seems plausible that preference reversals could be more common in other contexts, especially when information can be freely accessed.

Existing empirical work measures participants' *WTP* for a wide range of information structures. In Masatlioglu et al. (2022), participants are willing to pay on average \$0.08 to receive unavoidable, non-instrumental information (on the outcome of a lottery whose expected value is \$5) 30 minutes earlier. *WTP* for receiving positively skewed information early rather than negatively skewed information early is about three or four times higher. In Ganguly and Tasoff (2017), 83.4% of participants are willing to pay \$0.50 to affect the timing of information receipt in a similar context.

We believe it is notable that individuals appear to have a comparable *WTP* for committing not to see information (in our experiment) and seeing it sooner (in others), especially considering that commitments are implemented only probabilistically in our setting. Both behaviours are consistent with a model in which individuals are tempted by information, but they result in contradictory conclusions about preferences. Indeed, *WTP* positively predicts self-control costs at the individual level when commitments are not implemented: those who find it more difficult to resist information in session 2 are willing to pay *more* money to commit to avoiding it in session 1. In other settings, it cannot be ruled out that those who

are especially impatient to instantaneously acquire information when it is unavoidable are actually those with the strongest preference to avoid it ahead of time.

It has also been observed that individuals avoid undesired information in many settings in which it possibly has instrumental value (Eil and Rao, 2011; Oster et al., 2013). However, not all individuals avoid information. Our results raise the possibility that those individuals who do acquire information in such settings are simply those who are least able to exercise self-control.

### **5.3 Implications for policy**

Intrinsic preferences for information are of major policy relevance, especially in view of the dramatic growth of information markets in recent decades. Demonstrating that information is tempting raises a stronger case for managing the manufacture and creation of unwanted information. Our results demonstrate that many individuals strictly prefer to not be offered the information we generate, and are willing to pay both effort costs and money for that preference. In our setting, for many individuals, one interpretation is that the welfare-optimal outcome would have been for the information not to be created at all.

Social media is now well understood to be a temptation good whose welfare effects can be negative (Allcott et al., 2020; Allcott et al., 2022; Braghieri et al., 2022). Our findings suggest a deeper behavioural foundation for this phenomenon: one major function of social media is to supply information which may be tempting. Similar conclusions could be drawn for very recent developments in large language models, which promise an unprecedented ease of access to complex information.

Our treatment conditions also shed light on the potential role of notifications, now a fundamental feature of information supply on digital platforms. Our results suggest that the supply of partial information increases the strength of temptation to access undesirable information. This suggests the widespread use of notifications may be an effective strategy in eroding individuals' willingness to commit to reducing their use of social media, and that this could negatively impact their welfare. Further research could be informative on this and other related policy matters.

## References

- Alan, S., & Ertac, S. (2015). Patience, self-control and the demand for commitment: Evidence from a large-scale field experiment. *Journal of Economic Behavior & Organization*, *115*, 111–122.
- Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, *110*(3), 629–76.
- Allcott, H., Gentzkow, M., & Song, L. (2022). Digital addiction. *American Economic Review*, *112*(7), 2424–63.
- Augenblick, N., Niederle, M., & Sprenger, C. (2015). Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics*, *130*(3), 1067–1115.
- Bénabou, R., & Pycia, M. (2002). Dynamic inconsistency and self-control: A planner–doer interpretation. *Economics Letters*, *77*(3), 419–424.
- Bénabou, R., & Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, *117*(3), 871–915.
- Braghieri, L., Levy, R., & Makarin, A. (2022). Social media and mental health.
- Brocas, I., & Carrillo, J. D. (2008). The brain as a hierarchical organization. *American Economic Review*, *98*(4), 1312–46.
- Brunnermeier, M. K., & Parker, J. A. (2005). Optimal expectations. *American Economic Review*, *95*(4), 1092–1118.
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., et al. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.
- Caplin, A., & Leahy, J. (2001). Psychological expected utility theory and anticipatory feelings. *The Quarterly Journal of Economics*, *116*(1), 55–79.
- Caplin, A., & Leahy, J. (2004). The supply of information by a concerned expert. *The Economic Journal*, *114*(497), 487–505.
- Dekel, E., & Lipman, B. L. (2012). Costly self-control and random self-indulgence. *Econometrica*, *80*(3), 1271–1302.
- Eil, D., & Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, *3*(2), 114–38.
- Eliasz, K., & Schotter, A. (2007). Experimental testing of intrinsic preferences for noninstrumental information. *American Economic Review*, *97*(2), 166–169.
- Ely, J., Frankel, A., & Kamenica, E. (2015). Suspense and surprise. *Journal of Political Economy*, *123*(1), 215–260.



- Epstein, L. G. (2008). Living with risk. *The Review of Economic Studies*, 75(4), 1121–1141.
- Epstein, L. G., & Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57(4), 937–969.
- Falk, A., & Zimmermann, F. (2023). Attention and dread: Experimental evidence on preferences for information. *Management Science*.
- Fudenberg, D., & Levine, D. K. (2006). A dual-self model of impulse control. *American Economic Review*, 96(5), 1449–1476.
- Fudenberg, D., & Levine, D. K. (2012). Timing and self-control. *Econometrica*, 80(1), 1–42.
- Ganguly, A., & Tasoff, J. (2017). Fantasy and dread: The demand for information and the consumption utility of the future. *Management Science*, 63(12), 4037–4060.
- Golman, R., Hagmann, D., & Loewenstein, G. (2017). Information avoidance. *Journal of Economic Literature*, 55(1), 96–135.
- Golman, R., Loewenstein, G., Molnar, A., & Saccardo, S. (2021). The demand for, and avoidance of, information. *Management Science*.
- Grant, S., Kajii, A., & Polak, B. (1998). Intrinsic preference for information. *Journal of Economic Theory*, 83(2), 233–259.
- Gul, F., & Pesendorfer, W. (2001). Temptation and self-control. *Econometrica*, 69(6), 1403–1435.
- Heidhues, P., & Kőszegi, B. (2009). Futile attempts at self-control. *Journal of the European Economic Association*, 7(2-3), 423–434.
- Hilbert, M., & López, P. (2011). The world’s technological capacity to store, communicate, and compute information. *Science*, 332(6025), 60–65.
- Huffman, D., Raymond, C., & Shvets, J. (2022). Persistent overconfidence and biased memory: Evidence from managers. *American Economic Review*, 112(10), 3141–75.
- Kőszegi, B., & Rabin, M. (2009). Reference-dependent consumption plans. *American Economic Review*, 99(3), 909–36.
- Kreps, D. M., & Porteus, E. L. (1978). Temporal resolution of uncertainty and dynamic choice theory. *Econometrica*, 185–200.
- Masatlioglu, Y., Orhun, A. Y., & Raymond, C. (2022). Intrinsic information preferences and skewness. *Ross School of Business Paper*.
- Nielsen, K. (2020). Preferences for the resolution of uncertainty and the timing of information. *Journal of Economic Theory*, 189, 105090.
- Noor, J. (2007). Commitment and self-control. *Journal of Economic Theory*, 135(1), 1–34.
- O’Donoghue, T., & Rabin, M. (1999). Doing it now or later. *American Economic Review*, 89(1), 103–124.

- Oster, E., Shoulson, I., & Dorsey, E. (2013). Optimal expectations and limited medical testing: Evidence from Huntington disease. *American Economic Review*, *103*(2), 804–30.
- Roy-Chowdhury, V. (2022). Self-confidence and motivated memory loss: Evidence from schools.
- Royer, H., Stehr, M., & Sydnor, J. (2015). Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a fortune-500 company. *American Economic Journal: Applied Economics*, *7*(3), 51–84.
- Sacardo, S., & Serra-Garcia, M. (2023). Enabling or limiting cognitive flexibility? evidence of demand for moral commitment. *American Economic Review*, *113*(2), 396–429.
- Sadoff, S., Samek, A., & Sprenger, C. (2020). Dynamic inconsistency in food choice: Experimental evidence from two food deserts. *The Review of Economic Studies*, *87*(4), 1954–1988.
- Toussaert, S. (2018). Eliciting temptation and self-control through menu choices: A lab experiment. *Econometrica*, *86*(3), 859–889.
- Zimmermann, F. (2020). The dynamics of motivated beliefs. *American Economic Review*, *110*(2), 337–61.

## A1 Additional tables

Table A1: Implied binary information preferences from session 1

Binary preference	$N$	Share
$\{0\} \succ_1 \{1\}$	443	66.9% (1.8)
$\{0\} \sim_1 \{1\}$	122	18.4% (1.5)
$\{1\} \succ_1 \{0\}$	97	14.7% (1.4)
Total	662	100%

*Note:* Standard errors in parentheses, in percentage points.

Table A2: Menu preferences from session 1 — all stages

Menu preference	Type	Raw (no cost)		Main (effort cost)		$WTP > 0$	
		$N$	Share	$N$	Share	$N$	Share
$\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$	Strongly tempted	316	47.7% (1.9)	254	38.4% (1.9)	109	16.5% (1.4)
$\{0\} \sim_1 \{0, 1\} \sim_1 \{1\}$	Indifferent	13	2% (0.5)	80	12.1% (1.3)	144	21.8% (1.6)
$\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$	Strongly tempted	36	5.4% (0.9)	71	10.7% (1.2)	32	4.8% (0.8)
$\{0\} \sim_1 \{0, 1\} \succ_1 \{1\}$	Standard info. averse	20	3% (0.7)	54	8.2% (1.1)	198	29.9% (1.8)
$\{0, 1\} \succ_1 \{0\} \succ_1 \{1\}$	Flex	82	12.4% (1.3)	50	7.6% (1)	34	5.1% (0.9)
$\{0, 1\} \succ_1 \{1\} \succ_1 \{0\}$	Flex	55	8.3% (1.1)	34	5.1% (0.9)	28	4.2% (0.8)
$\{0, 1\} \succ_1 \{0\} \sim_1 \{1\}$	Flex	16	2.4% (0.6)	33	5% (0.8)	31	4.7% (0.8)
$\{1\} \succ_1 \{0, 1\} \succ_1 \{0\}$	Strongly tempted (info. loving)	51	7.7% (1)	29	4.4% (0.8)	21	3.2% (0.7)
$\{0, 1\} \sim_1 \{1\} \succ_1 \{0\}$	Other	13	2% (0.5)	17	2.6% (0.6)	28	4.2% (0.8)
$\{0\} \succ_1 \{1\} \succ_1 \{0, 1\}$	Flexibility averse	35	5.3% (0.9)	14	2.1% (0.6)	11	1.7% (0.5)
$\{1\} \succ_1 \{0\} \sim_1 \{0, 1\}$	Other	6	0.9% (0.4)	13	2% (0.5)	13	2% (0.5)
$\{0\} \sim_1 \{1\} \succ_1 \{0, 1\}$	Commitment loving	8	1.2% (0.4)	9	1.4% (0.5)	9	1.4% (0.5)
$\{1\} \succ_1 \{0\} \succ_1 \{0, 1\}$	Flexibility averse	11	1.7% (0.5)	4	0.6% (0.3)	4	0.6% (0.3)
Total		662	100%	662	100%	662	100%

*Note:* Participants must complete a short task for each strict preference submitted in their ranking. For  $WTP$  measure: if present, the first strict preference additionally requires participants to have a strictly positive willingness to pay to avoid a swap of their first and second choices. If  $WTP = 0$ , the first strict preference is replaced with indifference. Standard errors in parentheses, in percentage points.

Table A3: Menu preferences from session 1 — bonus takers only

Menu preference	Type	Raw (no cost)		Main (effort cost)		WTP > 0	
		N	Share	N	Share	N	Share
$\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$	Strongly tempted	289	43.7% (1.9)	231	34.9% (1.9)	96	14.5% (1.4)
$\{0\} \sim_1 \{0, 1\} \sim_1 \{1\}$	Indifferent	11	1.7% (0.5)	73	11% (1.2)	133	20.1% (1.6)
$\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$	Strongly tempted	31	4.7% (0.8)	64	9.7% (1.1)	28	4.2% (0.8)
$\{0, 1\} \succ_1 \{0\} \succ_1 \{1\}$	Flex	78	11.8% (1.3)	48	7.3% (1)	33	5% (0.8)
$\{0\} \sim_1 \{0, 1\} \succ_1 \{1\}$	Standard info. averse	17	2.6% (0.6)	48	7.3% (1)	182	27.5% (1.7)
$\{0, 1\} \succ_1 \{0\} \sim_1 \{1\}$	Flex	14	2.1% (0.6)	31	4.7% (0.8)	29	4.4% (0.8)
$\{0, 1\} \succ_1 \{1\} \succ_1 \{0\}$	Flex	49	7.4% (1)	29	4.4% (0.8)	24	3.6% (0.7)
$\{1\} \succ_1 \{0, 1\} \succ_1 \{0\}$	Strongly tempted (info. loving)	46	6.9% (1)	26	3.9% (0.8)	19	2.9% (0.6)
$\{0, 1\} \sim_1 \{1\} \succ_1 \{0\}$	Other	12	1.8% (0.5)	15	2.3% (0.6)	24	3.6% (0.7)
$\{1\} \succ_1 \{0\} \sim_1 \{0, 1\}$	Other	6	0.9% (0.4)	13	2% (0.5)	13	2% (0.5)
$\{0\} \succ_1 \{1\} \succ_1 \{0, 1\}$	Flexibility averse	30	4.5% (0.8)	12	1.8% (0.5)	9	1.4% (0.5)
$\{0\} \sim_1 \{1\} \succ_1 \{0, 1\}$	Commitment loving	8	1.2% (0.4)	8	1.2% (0.4)	8	1.2% (0.4)
$\{1\} \succ_1 \{0\} \succ_1 \{0, 1\}$	Flexibility averse	11	1.7% (0.5)	4	0.6% (0.3)	4	0.6% (0.3)
Total		602	100%	602	100%	602	100%

*Note:* Participants must complete a short task for each strict preference submitted in their ranking. For *WTP* measure: if present, the first strict preference additionally requires participants to have a strictly positive willingness to pay to avoid a swap of their first and second choices. If *WTP* = 0, the first strict preference is replaced with indifference. Standard errors in parentheses, in percentage points.

Table A4: Condensed types from session 1

Type	Unique rankings	Raw (no cost)		Main (effort cost)		WTP > 0	
		N	Share	N	Share	N	Share
<b>Strongly tempted</b>	1	316	47.7% (1.9)	<b>254</b>	<b>38.4%</b> (1.9)	109	16.5% (1.4)
Flexibility	3	153	23.1% (1.6)	117	17.7% (1.5)	93	14% (1.4)
Other	6	124	18.7% (1.5)	86	13% (1.3)	86	13% (1.3)
Indifferent	1	13	2% (0.5)	80	12.1% (1.3)	144	21.8% (1.6)
<b>Dynamic inconsistency</b>	1	36	5.4% (0.9)	<b>71</b>	<b>10.7%</b> (1.2)	32	4.8% (0.8)
Standard info. averse	1	20	3% (0.7)	54	8.2% (1.1)	198	29.9% (1.8)
Total	13	662	100%	662	100%	662	100%

*Note:* In main measure, participants must complete a short task for each strict preference submitted in their ranking. For *WTP* measure: if present, the first strict preference additionally requires participants to have a strictly positive willingness to pay to avoid a swap of their first and second choices. If *WTP* = 0, the first strict preference is replaced with indifference. Standard errors in parentheses, in percentage points.

Table A5: *WTP* by preference ranking

Ranking	Type	$E(WTP)$	% $WTP > 0$	$E(WTP WTP > 0)$	...of $N$
$\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$	Strongly tempted	\$0.09 (0.01)	42.9% (3.11)	\$0.2 (0.01)	254
$\{0\} \sim_1 \{0, 1\} \sim_1 \{1\}$	Indifferent	—	—	—	80
$\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$	Dynamic inconsistency	\$0.12 (0.02)	45.1% (5.95)	\$0.26 (0.03)	71
$\{0\} \sim_1 \{0, 1\} \succ_1 \{1\}$	Standard info. averse	\$0.2 (0.03)	67.9% (6.41)	\$0.29 (0.03)	54
$\{0, 1\} \succ_1 \{0\} \succ_1 \{1\}$	Flex	\$0.12 (0.02)	67.3% (6.7)	\$0.18 (0.02)	50
$\{0, 1\} \succ_1 \{1\} \succ_1 \{0\}$	Flex	\$0.19 (0.02)	82.4% (6.64)	\$0.23 (0.02)	34
$\{0, 1\} \succ_1 \{0\} \sim_1 \{1\}$	Flex	\$0.29 (0.03)	93.9% (4.22)	\$0.31 (0.03)	33
$\{1\} \succ_1 \{0, 1\} \succ_1 \{0\}$	Strongly tempted (info. loving)	\$0.23 (0.04)	72.4% (8.45)	\$0.32 (0.03)	29
$\{0, 1\} \sim_1 \{1\} \succ_1 \{0\}$	Other	\$0.27 (0.05)	82.4% (9.53)	\$0.32 (0.04)	17
$\{0\} \succ_1 \{1\} \succ_1 \{0, 1\}$	Flexibility averse	\$0.23 (0.05)	78.6% (11.38)	\$0.3 (0.05)	14
$\{1\} \succ_1 \{0\} \sim_1 \{0, 1\}$	Other	\$0.26 (0.05)	100% (0)	\$0.26 (0.05)	13
$\{0\} \sim_1 \{1\} \succ_1 \{0, 1\}$	Commitment loving	\$0.25 (0.07)	66.7% (16.67)	\$0.38 (0.04)	9
$\{1\} \succ_1 \{0\} \succ_1 \{0, 1\}$	Flexibility averse	\$0.15 (0.05)	100% (0)	\$0.15 (0.05)	4
Total					662

Note: Standard errors in parentheses, in \$ and percentage points as applicable.

Table A6: Logistic regression for preference categories

	$1(\{0\} \succ_1 \{1\})$	$1(\{0\} \succ_1 \{0, 1\}   \{0\} \succ_1 \{1\})$
	(1)	(2)
Age 25-34	-0.31 (0.30)	0.43 (0.34)
Age 35:44	-0.46 (0.32)	0.69 (0.37)
Age 45:54	-0.06 (0.37)	0.79 (0.43)
Age 55:64	-0.98 (0.42)	0.88 (0.55)
Age 65+	-0.65 (0.51)	1.06 (0.72)
Male	-0.05 (0.19)	0.04 (0.24)
Asian	-0.97 (0.29)	0.40 (0.44)
Black	-0.17 (0.39)	0.19 (0.51)
Mixed ethn.	-0.33 (0.37)	0.56 (0.50)
Other ethn.	-0.34 (0.41)	0.20 (0.55)
<i>PartAvoid</i>	-0.44 (0.18)	0.23 (0.23)
Risk tol.	-0.10 (0.04)	0.00 (0.05)
Altruism	-0.05 (0.04)	0.02 (0.05)
Patience	0.02 (0.05)	0.04 (0.06)
Strongly tempted	0.02 (0.04)	0.03 (0.06)
$p_a = 0.5$	0.07 (0.25)	-0.13 (0.33)
$p_a = 0.75$	0.35 (0.26)	0.06 (0.33)
$p_a = 1$	0.20 (0.26)	-0.05 (0.33)
Constant	1.93 (0.54)	-0.27 (0.63)
Observations	619	420
Log likelihood	-373.03	-242.70
Akaike inf. crit.	782.06	521.40

*Note:* Coefficients are log-odds.

Table A7: Measures of deliberation under exogenous  $\{0, 1\}$  in session 2 by session 1 preference ranking

Menu preference	Type	Time (seconds)	$N$
$\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$	Strongly tempted	13.1 (0.5)	105
$\{0\} \sim_1 \{0, 1\} \sim_1 \{1\}$	Indifferent	17.5 (1.9)	33
$\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$	Strongly tempted	18 (1.9)	28
$\{0\} \sim_1 \{0, 1\} \succ_1 \{1\}$	Standard info. averse	15.6 (2.3)	25
$\{0, 1\} \succ_1 \{0\} \succ_1 \{1\}$	Flex	21.5 (2.9)	19
$\{0, 1\} \succ_1 \{0\} \sim_1 \{1\}$	Flex	25.9 (3.2)	13
$\{0, 1\} \succ_1 \{1\} \succ_1 \{0\}$	Flex	16.6 (1.7)	10
$\{0, 1\} \sim_1 \{1\} \succ_1 \{0\}$	Other	11.7 (2.3)	7
$\{1\} \succ_1 \{0, 1\} \succ_1 \{0\}$	Strongly tempted (info. loving)	16.3 (3.4)	6
$\{0\} \succ_1 \{1\} \succ_1 \{0, 1\}$	Flexibility averse	11.5 (1.8)	6
$\{1\} \succ_1 \{0\} \sim_1 \{0, 1\}$	Other	17.3 (3.2)	5
$\{0\} \sim_1 \{1\} \succ_1 \{0, 1\}$	Commitment loving	12.5 (2.4)	3
$\{1\} \succ_1 \{0\} \succ_1 \{0, 1\}$	Flexibility averse	25.2 (11.8)	3
Total			263

Table A8: Probability of opening the envelope vs deliberation time under exogenous  $\{0\}$

$a = 1$	
(1)	
Time	3.13 (2.35)
(Time) <sup>2</sup>	-0.12 (0.09)
Constant	-22.84 (15.14)
Observations	133

*Note:* Coefficients are log-odds. Time spent on page where envelope decision is made, measured in seconds.

Table A9: Placebo test with session 1 time

	$\{0\} \succ_1 \{0, 1\}$
	$\log(WTP + 1)$
Not at all interested ( $m_1$ )	-0.30 (0.32)
Session 1 total time	-0.00 (0.00)
Session 1 total time <sup>2</sup>	0.00 (0.00)
Opened envelope ( $1 - \tilde{p}$ )	0.63 (0.80)
Constant	1.82 (0.65)
Observations	133

*Note:* Sample restricted to those randomised to face  $\{0, 1\}$  exogenously in session 2, those classified as tempted or strongly tempted in the main stage.

Table A10: Binary information preferences from session 1: treatment comparison

Binary preference	$N$		Share		Diff.	$p$ -value
	<i>PartAvoid</i>	<i>FullAvoid</i>	<i>PartAvoid</i>	<i>FullAvoid</i>		
$\{0\} \succ_1 \{1\}$	195	208	62.3% (2.7)	72% (2.6)	-9.67	0.01
$\{0\} \sim_1 \{1\}$	71	41	22.7% (2.4)	14.2% (2.1)	8.5	0.01
$\{1\} \succ_1 \{0\}$	47	40	15% (2)	13.8% (2)	1.18	0.77
Total	313	289	100%	100%		

*Note:* The 9% of participants who did not take the bonus are excluded from this table, since they were all allocated to the same condition (in which all of the information was accessed at the same time). Differences in percentage points.



Table A11: Interest in information by preference ranking and session, for participants choosing 0 from exogenous  $\{0, 1\}$  in session 2

Ranking	Type	Interest (s1)	Interest (s2)	$p$ -value	$N$
$\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$	Strongly tempted	1.3 (0.1)	1.5 (0.1)	0	103
$\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$	Strongly tempted	1.2 (0.1)	1.4 (0.2)	0.13	26
$\{0\} \sim_1 \{0, 1\} \sim_1 \{1\}$	Indifferent	1.8 (0.1)	2 (0.2)	0.06	26
$\{0\} \sim_1 \{0, 1\} \succ_1 \{1\}$	Standard info. averse	1.3 (0.1)	1.7 (0.2)	0.03	23
$\{0, 1\} \succ_1 \{0\} \succ_1 \{1\}$	Flex	2.2 (0.2)	2.3 (0.2)	0.43	17
$\{0, 1\} \succ_1 \{0\} \sim_1 \{1\}$	Flex	2.7 (0.2)	2.1 (0.3)	0.03	12
$\{0\} \succ_1 \{1\} \succ_1 \{0, 1\}$	Flexibility averse	1.6 (0.4)	1.6 (0.4)	—	5
$\{0, 1\} \sim_1 \{1\} \succ_1 \{0\}$	Other	1.8 (0.2)	2.5 (0.3)	0.22	4
$\{1\} \succ_1 \{0\} \sim_1 \{0, 1\}$	Other	3 (0.6)	2.3 (0.3)	0.18	3
$\{0, 1\} \succ_1 \{1\} \succ_1 \{0\}$	Flex	2.5 (1.5)	3 (0)	0.8	2
$\{1\} \succ_1 \{0\} \succ_1 \{0, 1\}$	Flexibility averse	2.5 (1.5)	2 (1)	0.5	2
$\{0\} \sim_1 \{1\} \succ_1 \{0, 1\}$	Commitment loving	1 (—)	2 (—)	—	1
$\{1\} \succ_1 \{0, 1\} \succ_1 \{0\}$	Strongly tempted (info. loving)	2 (—)	2 (—)	—	1
Total					225

*Note:*  $p$ -values correspond to two-sided paired  $t$ -tests of the difference in interest across sessions within each preference ranking. Responses to the question ‘How interested are you to know if the charity donation was reduced by \$15 as a result of your choice?’, numerically coded: ‘Not at all interested’ = 1; ‘A little interested’ = 2; ‘Fairly interested’ = 3; ‘Very interested’ = 4. Values are means, with standard errors in parentheses.

## A2 Experiment pages

Figure A1: Bonus choice page

All of your responses throughout this survey are fully anonymous and will not be analysed individually.

You can now accept an **additional \$4 bonus** for your participation in this study, at this stage increasing your total payment for this study by **200% to \$6**.

As part of this study, we will make a donation to an international charity on the behalf of a minority of participants. If you are one of these participants, taking the \$4 bonus means we have to reduce the donation by \$15 so we can meet our budget. For the majority of participants, however, accepting the bonus has no consequences at all.

Would you like to accept the \$4 bonus?

Accept the bonus	<input type="radio"/>
Turn the bonus down	<input type="radio"/>

## Figure A2: Charity information page

The donation will be made to Save the Children, an international charity focusing on children in the US and around the world. Donations are often used for purposes such as providing food and shelter to children in vulnerable situations.

### We champion the rights of the world's 2.3 billion children

In the U.S. and around the world, [Save the Children does whatever it takes](#) – every day and in times of crisis – to ensure children grow up healthy, educated and safe.

We are often the first or only child-focused organization working in the hardest-to-reach places, where it's toughest to be a child.

We will pay the charity after everyone has completed session 2, depending on how much the donation had to be reduced by.



## Figure A3: Preference elicitation — step 1 (*FullAvoid*)

You took the \$4 bonus. By default, in the **next session (in 1-2 days) we will generate a virtual envelope for you**, which you can choose to open. Inside, there is a message telling you there was either a **0% or 20% chance** that taking the bonus meant the donation to Save the Children was reduced by \$15. **Seeing what is inside the envelope will cost you \$0.25**, taken from your bonus pay.

What will the envelope look like next time?

If you open the envelope, you can then confirm if \$15 was actually removed from the charity donation. The next session will take about the same amount of time regardless of what you choose.

Please rank the following options so your **favorite has rank 1** and so on. The higher you rank an option, the more likely it is to actually happen. Feel free to give two or more options the same ranking. You will need to verify your ranking on the next page, so be mindful of how you choose.

	1	2	3
Automatically open the envelope next time.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do not generate the envelope at all.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ask me if I want to open the envelope next time.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure A4: Preference elicitation — step 1 (*PartAvoid*)

You took the \$4 bonus. At the start of the **next session (in 1-2 days)**, you will first be notified that there was either a 0% or 20% chance that taking the bonus meant the donation to Save the Children was reduced by \$15.

If you find out there was a 20% chance of the charity donation being reduced, by default **we will generate a virtual envelope for you**, which you can choose to open. Inside will be a message confirming whether \$15 was removed from the charity donation. The next session will take about the same amount of time regardless of what you choose. **Seeing what is inside the envelope will cost you \$0.50**, taken from your bonus pay.

What will the envelope look like next time?

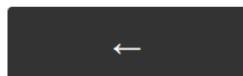
Please rank the following options so your **favorite has rank 1** and so on. The higher you rank an option, the more likely it is to actually happen. Feel free to give two or more options the same ranking. You will need to verify your ranking on the next page, so be mindful of how you choose.

	1	2	3
Do not generate the envelope at all.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ask me if I want to open the envelope next time.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Automatically open the envelope next time.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure A5: Preference elicitation — step 2

Please confirm your ranking. Note that the options may be displayed in a different order now.

	1	2	3
Ask me if I want to open the envelope next time.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Do not generate the envelope at all.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Automatically open the envelope next time.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>



### Figure A6: Preference elicitation — step 3

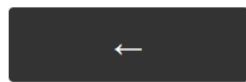
You ranked **not generating the envelope at all** above **deciding whether to open the envelope next time**.

If you definitely have a clear preference between these two options, we will shortly ask you to complete another brief sliders task to confirm it.

Do you definitely have a clear preference between these two options?

I definitely prefer to not generate the envelope at all

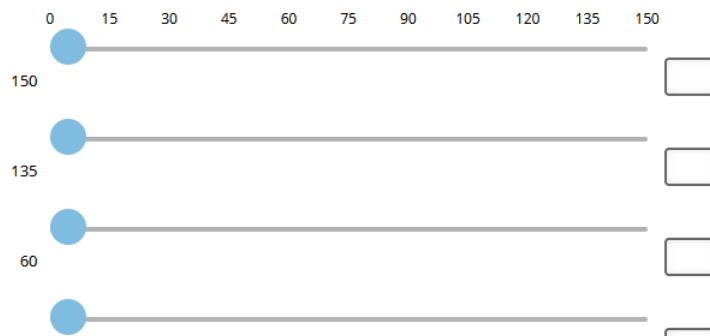
I don't have a clear preference



### Figure A7: Preference elicitation — effort confirmation

You must complete this page to confirm your preference for **not generating the envelope at all** over **deciding whether to open the envelope next time**. If you actually do not have a clear preference, feel free to navigate back and say so.

Please drag the sliders to the values indicated on the left.



## Figure A8: Preference elicitation — summary

As a reminder, the envelope contains a message allowing you to find out if the charity donation was reduced by \$15 as a result of your choice. By default, it will be generated when you complete the next session of this study, in 1-2 days. It would cost you \$0.25 of your bonus to open the envelope.

**Your best choice was not generating the envelope at all. Your second best was deciding whether to open the envelope next time. Your worst was automatically opening the envelope next time.**

If that's not right, please navigate back and change your ranking. Remember, the higher you rank a choice, the more likely it is to actually happen.

[More information](#)

*Optional:*

Can you tell us a bit about why you preferred not to generate the envelope rather than make your decision next time?

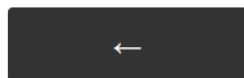


Figure A9: Preference elicitation — effort task

Please drag the sliders to the values indicated on the left.

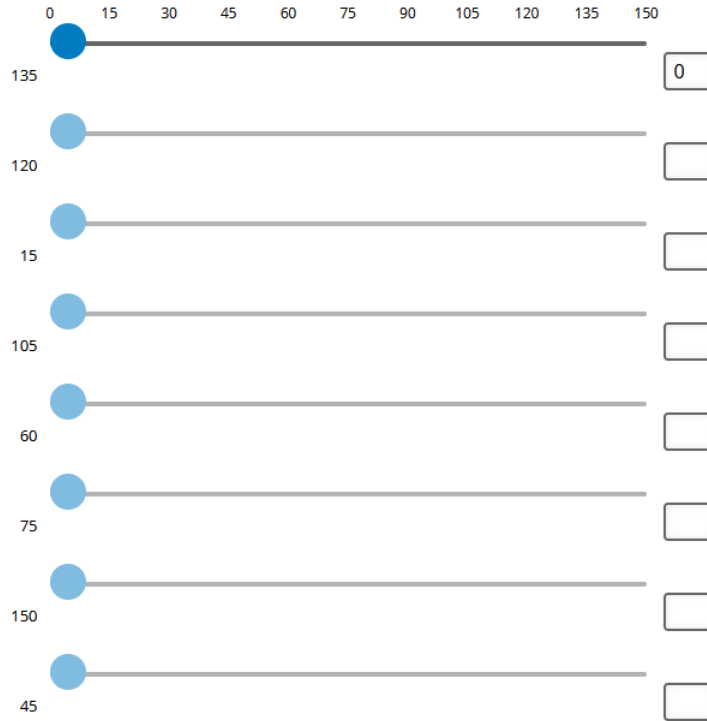


Figure A10: WTP elicitation

You said your best choice was to **not generate the envelope at all** and your next best was to **decide whether to open the envelope next time**. This question is about how much you value this preference in monetary terms.


What is the highest payment you'd be willing to forgo to **keep your best and next best choices in their order**? Please note this involves a randomly generated actual bonus payment between \$0 and \$0.50: if this payment is **higher** than your answer, then you will receive the payment and your choices will be switched, making your next best choice much more likely to happen.

0 5 10 15 20 25 cents 30 35 40 45 50

Highest payment you'd be willing to give up



Figure A11: Envelope — *FullAvoid*



Here is your virtual envelope. Inside is a message which will tell you that there was either a **0% or a 20% chance that \$15 was removed from the donation** pot as a result of your choice.


If it says there was a 20% chance of \$15 being removed, you will then be asked whether you want to confirm if this was the case.

Do you want to open the envelope now?

Figure A12: Envelope — *PartAvoid*


**NOTIFICATION B**

There is a **20% chance** that \$15 has been taken from the donation to Save the Children as a result of you taking the bonus.



Here is your virtual envelope. Inside, there is a message telling you for sure if money has been taken from the charity donation as a result of your choice.

Figure A13: Envelope — bonus rejected



Here is your virtual envelope. Inside is a message which will tell you whether you could have taken the bonus without affecting the donation to Save the Children.



### A3 Why is information tempting?

Section 2 outlined a generic model of temptation and self-control in order to introduce the main predictions of our experiment and precisely characterise the preference rankings we observe. Here, we embed a representation of intrinsic information preferences within that model in order to help to rationalise our results, including the difference we observe between treatment arms.

Our task is to specify the commitment and temptation utilities  $g_1(a)$  and  $g_2(a)$  as a function of information. Recall that  $a$  denotes a binary action to access information. We now define the random variable  $\theta \in \{0, 1\}$  as the event that the individual was not a *Donor*, focusing on the case where the individual took the \$4 payment for simplicity.

Based on the large set of free-text responses, many participants in our experiment who commit to avoiding information appear to engage in wishful thinking about their chance of not being a *Donor*. We make this feature central to our formalisation of information preferences, since it also easily rationalises our treatment effect. We introduce the variable  $\theta(f(\Omega(a)))$ , which captures the individual's *wishful* perception of the value of  $E(\theta)$  with the information set  $\Omega$  conditional on the action  $a$  to obtain information. This value can differ from the rational expectation of  $\theta$  given the individual's information set. However, the individual is constrained in setting  $\theta$  by the rational belief distribution over  $\theta$ ,  $f(\Omega(a))$ . Suppose after setting  $a$  and receiving the information set  $\Omega(a)$ ,  $\theta$  maximises the following function for  $\tilde{\theta}$  for  $\tilde{\theta} \in [0, 1]$ , with  $\kappa_1 > 0$ :

$$\tilde{\theta} - \kappa_1 \frac{(\tilde{\theta} - E_{f(\Omega(a))}(\theta))^2}{Var_{f(\Omega(a))}(\theta)}. \quad (5)$$

The individual trades off two considerations in deciding on  $\theta$ . The first is a desire to maintain a positive state of mind about the consequence of a selfish action, captured by the first term. The second is a desire to minimise cognitive dissonance, in that deviations of  $\theta$  from the rational subjective distribution are penalised. Suppose for now that the action  $a$  involves fully resolving uncertainty about  $\theta$ , setting  $a = 1$  costs  $p_a$  and the individual values money according to the function  $\omega(\cdot)$ . Then, we can write commitment utility as

$$g_1(a) = E \left( \theta(f(\Omega(a))) - \kappa_1 \frac{(\theta(f(\Omega(a))) - E_{f(\Omega(a))}(\theta))^2}{Var_{f(\Omega(a))}(\theta)} - \omega(p_a a) \right). \quad (6)$$

Note that if all uncertainty is resolved,  $Var_{f(\Omega(a))} = 0$  and (5) implies  $\theta = \tilde{\theta}$ ; the individual cannot engage in wishful thinking when they know the truth. Thus, the expected commitment

payoff to setting  $a = 1$  is simply

$$g_1(1) = E_\mu(\theta) - \omega(p_a a). \quad (7)$$

On the other hand, the commitment payoff to setting  $a = 0$  involves sustaining the prior distribution function  $f(\theta; \Omega(0)) = \mu$ , so

$$g_1(0) = \theta(\mu) - \kappa_1 \frac{(\theta(\mu) - E_\mu(\theta))^2}{Var_\mu(\theta)}. \quad (8)$$

Note that  $\theta(\mu) > E_\mu(\theta)$ : the marginal effect of increasing  $\tilde{\theta}$  in (5) is strictly positive at  $\tilde{\theta} = E_\mu(\theta)$ . As such, we know for sure that  $g_1(0) < g_1(1)$ :  $\theta = E_\mu(\theta)$  is not the optimal solution for (5), meaning commitment utility is unambiguously reduced by accessing information even before taking into account the cost of accessing information  $p_a a$ . This establishes the motive for individuals to want to avoid information *ex ante*.

It now remains to establish why information could be *tempting* in period 2. In order to do so, our temptation utility function,  $g_2(a)$ , must have a stronger preference for information than  $g_1(a)$ . Guided by our result that subjective interest in acquiring the information increases in period 2, and ‘curiosity’ is commonly cited as a reason for reversing period 1 preferences to access information, we posit that  $g_2(a)$  has the following form, with the parameter  $\kappa_2 > 0$ ,

$$g_2(a) = \theta(f(\Omega(a))) - \kappa_1 \frac{(\theta(f(\Omega(a))) - E_{f(\Omega(a))}(\theta))^2}{Var_{f(\Omega(a))}(\theta)} - \kappa_2 Var_{f(\Omega(a))}(\theta) - \omega(p_a a). \quad (9)$$

$g_2(a)$  thus includes an extra term,  $\kappa_2 Var_{f(\Omega(a))}(\theta)$ , capturing heightened curiosity about the truth: the individual becomes purely averse to uncertainty in the rational subjective distribution on  $\theta$  at the time of information becoming available. It is easy to see that since  $a$  is informative for  $\theta$ ,  $g_2(1) > g_1(1)$ . Thus, all individuals are strictly more attracted to information under temptation utility than commitment utility. However, information is more likely to be *tempting* for those with relatively low  $\kappa_1$  and high  $\kappa_2$ . These individuals find it relatively easy to engage in wishful thinking in the absence of information, but anticipate being curious about the truth when it is accessible. This curiosity motive either makes it more costly to refuse information when it is offered (if they successfully exert self-control), or makes them more likely to obtain information (if they fail to exert self-control).

This model of intrinsic information preferences also puts us in a position to interpret the difference between the treatment arms *FullAvoid* and *PartAvoid*. The key difference between them is that in *FullAvoid*, avoiding information involves maintaining a more diffuse belief on  $\theta$  than in *PartAvoid*, where the signal  $s = s_2$  must have been observed prior to

setting  $a$ . Suppose the rational subjective distribution after observing  $s_2$  is  $f(\Omega(0)) = \mu_2$ . We know that  $E_{\mu_2}(\theta) = 0.8 < E_{\mu}(\theta)$  and  $Var_{\mu_2}(\theta) < Var_{\mu}(\theta)$ . Thus,  $\theta(\mu_2) < \theta(\mu)$  and  $g_1(0)$  is unambiguously lower under in *PartAvoid* than in *FullAvoid*. This means the incentive to set  $a = 0$  is attenuated under commitment utility in *PartAvoid*: since an unfavourable signal has already been observed in the case where information is offered, the individual is less able to engage in wishful thinking. This matches our results in a few attractive ways. The prevalence of tempted types is equal across the two treatment groups, but the prevalence of strongly tempted types is lower in *PartAvoid*. These types expect relatively high self-control costs, so  $\kappa_2$  is large. The analysis above suggests that being in condition *PartAvoid* nudges these types away from submitting the strict preference  $\{0\} \succ_1 \{0, 1\}$  because the commitment utility evaluation of  $a = 0$  is worse: individuals know they will be less able to engage in wishful thinking when they avoid information.

Table A12: Happiness with \$4 choice by preference ranking

Ranking	Type	Happiness with \$4 choice	$N$
$\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$	<b>Strongly tempted</b>	<b>4.1 (0.1)</b>	254
$\{0\} \sim_1 \{0, 1\} \sim_1 \{1\}$	Indifferent	3.9 (0.1)	80
$\{0\} \succ_1 \{0, 1\} \sim_1 \{1\}$	<b>Strongly tempted</b>	<b>3.8 (0.1)</b>	71
$\{0\} \sim_1 \{0, 1\} \succ_1 \{1\}$	<b>Standard info. averse</b>	<b>3.8 (0.1)</b>	54
$\{0, 1\} \succ_1 \{0\} \succ_1 \{1\}$	Flex	4.1 (0.1)	50
$\{0, 1\} \succ_1 \{1\} \succ_1 \{0\}$	Flex	3.8 (0.2)	34
$\{0, 1\} \succ_1 \{0\} \sim_1 \{1\}$	Flex	3.6 (0.1)	33
$\{1\} \succ_1 \{0, 1\} \succ_1 \{0\}$	Strongly tempted (info. loving)	4.3 (0.2)	29
$\{0, 1\} \sim_1 \{1\} \succ_1 \{0\}$	Other	3.6 (0.3)	17
$\{0\} \succ_1 \{1\} \succ_1 \{0, 1\}$	Flexibility averse	4.1 (0.2)	14
$\{1\} \succ_1 \{0\} \sim_1 \{0, 1\}$	Other	4.3 (0.3)	13
$\{0\} \sim_1 \{1\} \succ_1 \{0, 1\}$	Commitment loving	4.4 (0.3)	9
$\{1\} \succ_1 \{0\} \succ_1 \{0, 1\}$	Flexibility averse	4.2 (0.2)	4
Total			662

*Note:* Responses to the question ‘How happy are you with your decision to take the \$4 bonus?’, numerically coded: ‘Extremely unhappy’ = 1; ‘Somewhat unhappy’ = 2; ‘Neither happy nor unhappy’ = 3; ‘Somewhat happy’ = 4; ‘Extremely happy’ = 4. Means, with standard errors in parentheses.

We conclude this section with a final piece of empirical evidence. The model proposed in this section indicates wishful thinking crucially determines the motive to avoid information: if the individual is more able to hold optimistic beliefs about  $\theta$  in the absence of information, they are more able to submit the strict preference  $\{0\} \succ_1 \{0, 1\}$ . Recalling Proposition 1, tempted types are much more likely to have very strong commitment utility preferences for avoiding information, whereas this motive cannot be as strong for weak self-control and

standard information averse types. In the model set out above, this should imply tempted types are relatively able to engage in wishful thinking about the value of  $\theta$  in the absence of information. Table A12 reports individuals' subjective happiness with their \$4 bonus choice at the end of session 1. The data match well the predictions of our model: tempted types are relatively happy with their choice on the bonus compared to weak self-control and standard information averse types. Separately, flexibility types, who are less sure *ex ante* about avoiding information, are generally more negatively preoccupied with their choice.