Online Appendix

A Mathematical appendix

A.1 Additional Technical Results

We first introduce an additional lemma that characterizes the differentiability and continuity properties of the statistics employed in Theorem 1. The proof is given in Appendix A.3.

Lemma A.1. *Assume that individuals choose attention strategies optimally. Then,*

1. $\bar{W}(r)$ is differentiable almost everywhere.

   (a) $Pr(z=1|j,r)$ is increasing in r and differentiable almost everywhere.

   (b) $\bar{W}(r)$ is differentiable at any point r where $Pr(z=1|j=1,r)-Pr(z=1|j=0,r)$ is continuous in r.

   (c) Suppose that $\bar{K}_{ai}$ and $\bar{K}_{oi}$ are strictly convex for all i. Then, $\bar{W}(r)$ is everywhere continuously differentiable.

   (d) Suppose that at (p,r), $Pr(j=1|p,r)$ is continuously differentiable in p and that $\bar{W}(r)$ is continuously differentiable in r. Then $Pr(j=1|p,r)$ is continuously differentiable in r and $Pr(z=1|p,r)$ is continuously differentiable in p.

Lemma A.1 allows us to express some of our main results in terms of marginal conditions without much loss of generality. Parts 1 and 2 of the lemma show that, without any additional assumptions, two of the key statistics are differentiable everywhere except on a set of Lebesgue measure zero. Part 3 of the lemma concerns the condition that $Pr(z=1|j=1,r)-Pr(z=1|j=0,r)$ is continuous in r. This is a plausible condition in our experiments, where we find that $Pr(z=1|j,r)$ for $j\in\{0,1\}$ changes negligibly when we increase r by a small amount. Part 4 provides an alternative set of assumptions for differentiability of $\bar{W}(r)$, which is that the cost functions are convex. Finally, part 5 considers the mild assumption that $Pr(j=1|p,r)$ is differentiable in p. This is a natural condition on a demand function for BEs and holds whenever the distribution of individual differences is smooth. For example, this condition holds when $\bar{K}^1_{ai}(q)-\bar{K}^0_{ai}(q)=\bar{K}^1_{a}(q)-\bar{K}^0_{a}(q)+\eta_i$, where η_i is a random variable with a smooth density function that is interpreted as a person-specific nuisance cost of the BE.

A.2 Preliminaries for Proofs of Main Results

A.2.1 Notation

By reasoning analogous to that of Lemma 1, we can express the indirect utility functions as

$$V_i^j(r) = \max_{q\in[\underline{q},\bar{q}]} \left\{ rq - \bar{K}_i^j(q) \right\}$$

where $\underline{q} = \underline{q}_a \underline{q}_o$, $\bar{q} = \bar{q}_a \bar{q}_o$, and $\bar{K}_i^j(q) = \inf_{q_a,q_o}\{K_{ai}^j(q_a) + K_{oi}(q_o) | q_a q_o \geq q\}$.

Define the functions $f_i^j(r,q) = rq - \bar{K}^j(q)$, so that $V_i^j(r) = \max_q f_i^j(r,q)$. Define $X_i^j(r) = \{q | f_i^j(r,q) = V_i^j(q)\}$ as the maximizers of f_i^j, and note that by assumption X_i^j is non-empty. Under the assumption of optimality, an individual's choice of q under technology j is a selection $q_i^j(r)$ from $X_i^j(r)$.

Define $\mathscr{V}_i(p,r) = \max\{V_i^1(r) - p, V_i^0(r)\}$, and define $\bar{\mathscr{V}}(p,r) = \mathbb{E}_i \mathscr{V}_i(p,r)$. We can write $\mathscr{V}_i(p,r) = \max_{q,j} \varphi_i(q,j,p,r)$ where

$$\varphi_i = j\left(rq - \bar{K}_i^1(q) - p\right) + (1-j)\left(rq - \bar{K}_i^0(q)\right).$$

Similarly, define $Y_i(p,r) = \{(q,j) | \varphi_i(q,j,p,r) = \mathscr{V}_i(p,r)\}$ as the maximizers of φ_i, which again is non-empty by assumption. An individual's choice of technology and completion probability is a selection $(j_i(p,r), q_i(r)) \in Y_i(p,r)$. We define $Pr_i(z=1|p,r)$ as individual i's probability of successfully completing the task, given by $j_i(p,r)q_i^1(r) + (1 - j_i(p,r))q_i^0(r)$.

A.2.2 Preliminary Lemmas

Lemma A.2. *$V_i^j(r)$ is strictly increasing in r. Any selection $q_i^j(r)$ is increasing in r.*

Proof. Consider $r_2 > r_1$. Then

$$\begin{aligned} V_i^j(r_2) &\geq f_i^j(r_2, q_i^j(r_1)) \\ &> f_i^j(r_1, q_i^j(r_1)) \\ &= V_i^j(r_1) \end{aligned}$$

which establishes the first claim. Next,

$$
\begin{aligned}
& r_2 q_i^j(r_2) - \bar{K}_i^j(q_i^j(r_2)) \geq r_2 q_i^j(r_1) - \bar{K}_i^j(q_i^j(r_1)) \\
\Leftrightarrow\ & r_2 \left(q_i^j(r_2) - q_i^j(r_1) \right) \geq \bar{K}_i^j(q_i^j(r_2)) - \bar{K}_i^j(q_i^j(r_1)).
\end{aligned} \tag{1}
$$

Similarly,

$$
r_1 \left(q_i^j(r_1) - q_i^j(r_2) \right) \geq \bar{K}_i^j(q_i^j(r_1)) - \bar{K}_i^j(q_i^j(r_2)). \tag{2}
$$

Combining (1) and (2) implies that

$$
\begin{aligned}
r_2 \left(q_i^j(r_2) - q_i^j(r_1) \right) & \geq \bar{K}_i^j(q_i^j(r_2)) - \bar{K}_i^j(q_i^j(r_1)) \\
& \geq r_1 \left(q_i^j(r_2) - q_i^j(r_1) \right).
\end{aligned}
$$

Since $r_2 > r_1$, the above equality can only hold if $q_i^j(r_2) - q_i^j(r_1)$ is non-negative, which establishes the second part of the claim. □

Lemma A.3. *If $Pr(z=1|p,r)$ is continuous r, then $\bar{\mathscr{V}}$ is differentiable in r. If $Pr(j=1|p,r)$ is continuous in p, then $\bar{\mathscr{V}}$ is differentiable in p.*

Proof. Define $x = ((j_i, q_i))_{i \in \mathscr{I}}$ as the tuple of strategies of all individuals $i \in \mathscr{I}$ in the data. Define $\varphi(x,p,r) = \mathbb{E}_i \varphi_i(j_i, q_i, p, r)$, and note that x is a maximizer of φ if (j_i, q_i) is a maximizer of φ_i for each i. Thus, $\bar{\mathscr{V}}(p,r) = \max_x \varphi(x,p,r)$. Now, because $\varphi(x,p,r)$ is linear in r and p, and because $\frac{\partial}{\partial r}\varphi(x,p,r)$ and $-\frac{\partial}{\partial p}\varphi(x,p,r)$ are contained in the unit interval, all assumptions of Theorem 3 of Milgrom and Segal (2002) are satisfied. Thus, $\bar{\mathscr{V}}(p,r)$ is left- and right-hand differentiable in both r and p, with the respective derivatives given by

$$
\begin{aligned}
\frac{d_-}{dr}\bar{\mathscr{V}}(p,r) &= \lim_{x\to r-} \mathbb{E}_i Pr_i(z=1|p,r) \\
&= \lim_{x\to r-} Pr(z=1|p,r) \\
\frac{d_+}{dr}\bar{\mathscr{V}}(p,r) &= \lim_{x\to r+} \mathbb{E}_i Pr_i(z=1|p,r) \\
&= \lim_{x\to r+} Pr(z=1|p,r) \\
\frac{d_-}{dp}\bar{\mathscr{V}}(p,r) &= \lim_{x\to p-} \mathbb{E}_i j_i(p,r)(-1) \\
&= \lim_{x\to p-} -Pr(j=1|p,r) \\
\frac{d_+}{dp}\bar{\mathscr{V}}(p,r) &= \lim_{x\to p+} \mathbb{E}_i j_i(p,r)(-1) \\
&= \lim_{x\to p+} -Pr(j=1|p,r).
\end{aligned}
$$

When $Pr(z=1|p,r)$ is continuous in r, the left and right limits are equal, and thus $\bar{\mathscr{V}}(p,r)$ is differentiable in r. Similarly, $\bar{\mathscr{V}}(p,r)$ is differentiable in p when $Pr(j=1|p,r)$ is continuous in p. □

A.3 Proofs of Main Results

Proof of Lemma 1

Proof. Suppose first that (s_a^*, s_o^*) is a solution to (1), and define $q_a^* = \mathbb{E}Q(s_a^*, \omega_a)$ and $q_o^* = \mathbb{E}Q(s_o^*, \omega_o)$. An individual maximizing (2) can achieve at least

$$\mathbb{E}\left[rQ_a(s_a^*, \omega_a)Q_o(s_o^*, \omega_o) - K_{ai}(s_a^*) - K_{oi}(s_o^*, \omega_o)\right]$$

by setting $q_a = q_a^*$ and $q_o = q_o^*$. We now show that the individual cannot do any better. By way of contradiction, assume that there exist (q_a', q_o') such that

$$rq_a'q_o' - \bar{K}_{ai}(q_a') - \bar{K}_{oi}(q_o') \geq \mathbb{E}\left[rQ_a(s_a^*, \omega_a)Q_o(s_o^*, \omega_o) - K_{ai}(s_a^*) - K_{oi}(s_o^*, \omega_o)\right] + \varepsilon$$

for some $\varepsilon > 0$. By definition of the $\bar{K}$ functions, there exist (s_a', s_o') such that $\mathbb{E}Q_a(s_a', \omega_a) \geq q_a'$, $\mathbb{E}Q_o(s_o', \omega_a) \geq q_o'$ and $K_{ai}(s_a') \leq \bar{K}_{ai}(q_a') + \varepsilon/4$, $\mathbb{E}K_{oi}(s_o', \omega_o) \leq \bar{K}_{oi}(q_o') + \varepsilon/4$. Thus,

$$\mathbb{E}\left[rQ_a(s'_a,\omega_a)Q_o(s'_o,\omega_o)-K_{ai}(s'_a)-K_{oi}(s'_o,\omega_o)\right]\geq\mathbb{E}\left[rQ_a(s^*_a,\omega_a)Q_o(s^*_o,\omega_o)-K_{ai}(s^*_a)-K_{oi}(s^*_o,\omega_o)\right]+\varepsilon/2$$

which contradicts the optimality of (s^*_a,s^*_o).

To prove the converse direction, note again that by definition of the $\bar{K}$ functions, for any $\varepsilon>0$, there exist (s^*_a,s^*_o) such that $\mathbb{E}Q_a(s^*_a,\omega_a)\geq q^*_a$, $\mathbb{E}Q_o(s'_o,\omega_a)\geq q^*_o$ and $K_{ai}(s^*_a)\leq\bar{K}_{ai}(q'_a)+\varepsilon/2$, $K_{oi}(s^*_o)\leq\bar{K}_{oi}(q^*_o)+\varepsilon/2$. Thus,

$$\max_{(s_a,s_o)\in S_a\times S_o}\mathbb{E}\left[rQ_a(s_a,\omega_a)Q_o(s_o,\omega_o)-K_{ai}(s_a)-K_{oi}(s_o,\omega_o)\right]\geq rq^*_aq^*_o-\bar{K}_{ai}(q^*_a)-\bar{K}_{oi}(q^*_o)-\varepsilon$$

for any $\varepsilon>0$, and thus

$$\max_{(s_a,s_o)\in S_a\times S_o}\mathbb{E}\left[rQ_a(s_a,\omega_a)Q_o(s_o,\omega_o)-K_{ai}(s_a)-K_{oi}(s_o,\omega_o)\right]\geq rq^*_aq^*_o-\bar{K}_{ai}(q^*_a)-\bar{K}_{oi}(q^*_o).$$

On the other hand, as we have already argued in the first part of the proof, the agent cannot find strategies (s_a,s_o) that obtain higher expected utility than $rq^*_aq^*_o-\bar{K}_{ai}(q^*_a)-\bar{K}_{oi}(q^*_o)$. □

Proof of Lemma 2

Proof. We need to show that for any q_1,q_2 and $\alpha\in(0,1)$,

$$\bar{K}_a(\alpha q_1+(1-\alpha)q_2)<\alpha\bar{K}_a(q_1)+(1-\alpha)\bar{K}_a(q_2).$$

The argument for $\bar{K}_o$ is identical. By convexity of S_a, for any $q\in[\underline{q}_a,\bar{q}_a]$ there must be some $s\in S_a$ such that $\mathbb{E}Q(s,\omega_a)=q$. Thus, we can choose $s_1,s_2\in S_a$ such that $s_1\in\text{argmin}_s\{K_a(s)|\mathbb{E}Q(s,\omega_a)\geq q_1\}$ and analogously for s_2. Because $\mathbb{E}Q(\cdot,\omega_a)$ is concave, we have

$$\begin{aligned}\bar{K}_a(\alpha q_1+(1-\alpha)q_2)&\leq\bar{K}_a(\mathbb{E}Q(\alpha s_1+(1-\alpha)s_2,\omega_a)) &(3)\\&\leq K_a(\alpha s_1+(1-\alpha)s_2) &(4)\\&\leq\alpha K_a(s_1)+(1-\alpha)K_a(s_2) &(5)\\&=\alpha\bar{K}_a\left(\mathbb{E}Q(s_1,\omega_a)\right)+(1-\alpha)\bar{K}_a\left(\mathbb{E}Q(s_1,\omega_a)\right). &(6)\end{aligned}$$

Line (3) follows by the concavity of $\mathbb{E}Q(\cdot,\omega_a)$. Line (4) follows by the definition of $\bar{K}_a$. Line (5) follows by convexity of K_a, and line (6) follows by the definition of s_1 and s_2. $\square$

Proof of Lemma A.1

Proof. Part 1: In Lemma A.1, we have shown that $V_i^j(r)$ is strictly increasing in r. Thus, $\mathbb{E}_i V_i^j(r)$ is strictly increasing in r, and differentiable almost everywhere. It follows that $\bar{W}(r) = \mathbb{E}_i V_i^1(r) - \mathbb{E}_i V_i^0(r)$ is differentiable almost everywhere.

Part 2: We can write

$$\varphi_i(j,q,p,r) = rq - \psi_i(j,q,p,r)$$

where $\psi_i(j,q,p) = jp + j\left(\bar{K}_i^1(q) - \bar{K}_i^0(q)\right) + \bar{K}_i^0(q)$.

Consider $r_2 > r_1$. Then, the optimal selections $j_i(r)$ and $q_i(r)$ satisfy

$$\begin{aligned} r_2 q_i(r_2) - \psi_i(j_i(r_2), q_i(r_2), p) &\geq r_2 q_i(r_1) - \psi_i(j_i(r_1), q_i(r_1), p) \\ \Leftrightarrow r_2\left(q_i(r_2) - q_i(r_1)\right) &\geq \psi_i(j_i(r_2), q_i(r_2), p) - \psi_i(j_i(r_1), q_i(r_1), p). \end{aligned}$$

Similarly,

$$r_1\left(q_i(r_1) - q_i(r_2)\right) \geq \psi_i(j_i(r_1), q_i(r_1), p) - \psi_i(j_i(r_2), q_i(r_2), p)$$

and thus

$$r_2\left(q_i(r_2) - q_i(r_1)\right) \geq r_1\left(q_i(r_2) - q_i(r_1)\right),$$

which can hold only if $q_i(r_2) - q_i(r_1) \geq 0$. Thus, $Pr_i(j=1|p,r)$ is increasing in r for all i, and therefore $Pr(j=1|p,r)$ is increasing in r as well. The monotonicity implies almost everywhere differentiability.

Part 3: Define $x^j = (q_i^j)_{i\in\mathscr{I}}$ as the tuple of strategies of all individuals $i \in \mathscr{I}$ in the data given technology j. Define $f^j(x^j,r) = \mathbb{E}_i f_i^j(q_i,r)$, and note that x^j is a maximizer of f if q_i^j is a maximizer of f_i^j for each i. Thus, $\bar{W}(r) = \max_{x^1} f^1(x^1,r) - \max_{x^0} f^0(x^0,r)$. Now because $f^j(x,p,r)$ is linear in r, and because $\frac{\partial}{\partial r} f^j(x,p,r)$ is contained in the unit interval, all assumptions of Theorem 3 of Milgrom and Segal (2002) are satisfied. Thus, $\bar{W}(r)$ is left-and right-hand differentiable in r, with the respective derivatives given by

$$\begin{aligned}
\frac{d_{-}}{dr}\bar{W}(p,r) &= \lim_{x\to r-}\left(\mathbb{E}_i Pr_i(z=1|j=1,r)-\mathbb{E}_i Pr_i(z=1|j=0,r)\right)\\
&= \lim_{x\to r-} D(z=1|p,r)\\
\frac{d_{+}}{dr}\bar{W}(p,r) &= \lim_{x\to r+}\left(\mathbb{E}_i Pr_i(z=1|j=1,r)-\mathbb{E}_i Pr_i(z=1|j=0,r)\right)\\
&= \lim_{x\to r+} D(z=1|p,r).
\end{aligned}$$

When $D(z=1|r)$ is continuous in r, the left- and right-hand limits are equal, and thus $\bar{W}(r)$ is continuously differentiable in r.

Part 4: If $\bar{K}_{ai}$ and $\bar{K}_{oi}$ are strictly convex for all i, then $\bar{K}_i^j$ (as defined in Appendix A.2.1) is strictly convex, by an argument identical to that in the proof of Lemma 2. Thus, each individual's optimal choice $q_i^j(r)$ is unique for each (j,r). Moreover, since convex functions are continuous, this implies that f_i^j is continuous. Thus, Corollary 4 of Milgrom and Segal (2002) implies that $V_i^j(r)$ is everywhere differentiable in r, with derivative $q_i^j(r)$. The claim then follows immediately.

Part 5: Note that $Pr(j=1|p,r)=Pr\left(V_i^1(r)-V_i^0(r)-p\geq 0\right)$. Now, since $V_i^j(r)$ is increasing, it has left- and right-hand derivatives everywhere. Thus, $Pr\left(V_i^1(r)-V_i^0(r)-p\geq 0\right)$ is left- and right-hand differentiable everywhere. Now, by the assumption that $Pr(j=1|p,r)$ is continuously differentiable in p,

$$\begin{aligned}
\frac{d-}{dr}Pr\left(V_i^1(r)-V_i^0(r)-p\geq 0\right) &= \frac{d}{dp}Pr\left(V_i^1(r)-V_i^0(r)-p\geq 0\right)\frac{d-}{dr}\mathbb{E}_i\left(V_i^1(r)-V_i^0(r)\right)\\
&= \frac{d}{dp}Pr\left(V_i^1(r)-V_i^0(r)-p\geq 0\right)\frac{d-}{dr}\bar{W}(r)\\
\frac{d+}{dr}Pr\left(V_i^1(r)-V_i^0(r)-p\geq 0\right) &= \frac{d}{dp}Pr\left(V_i^1(r)-V_i^0(r)-p\geq 0\right)\frac{d+}{dr}\mathbb{E}_i\left(V_i^1(r)-V_i^0(r)\right)\\
&= \frac{d}{dp}Pr\left(V_i^1(r)-V_i^0(r)-p\geq 0\right)\frac{d+}{dr}\bar{W}(r).
\end{aligned}$$

Thus, $Pr\left(V_i^1(r)-V_i^0(r)-p\geq 0\right)$ is continuously differentiable in r if $\bar{W}(r)$ is continuously differentiable in r.

Next, to show that $Pr(z=1|p,r)$ is continuously differentiable in p, note that it is given by

$$Pr(j=1|p,r)Pr(z=1|j=1,r)+(1-Pr(j=1|p,r))Pr(z=1|j=0,r).$$

Thus, $Pr(z=1|p,r)$ is continuously differentiable in p if $Pr(j=1|p,r)$ is continuously differentiable in p. □

Proof of Theorem 1

Proof. Since f_i^j is a linear function of r, all assumptions of Theorem 2 of Milgrom and Segal (2002) are satisfied for f_i^j. Moreover, note that $\frac{\partial}{\partial r} f_i^j = q$. Thus, if $q_i^j(r)$ is an individual's optimal choice under technology j, we have that

$$V_i^j(r+\Delta)-V_i^j(r)=\int_{x=r}^{x=r+\Delta} q_i^j(r)dr.$$

Now

$$\begin{aligned}
\bar{W}(r+\Delta)-\bar{W}(r) &= \mathbb{E}_i\left[V_i^1(r+\Delta)-V_i^0(r+\Delta)\right]-\mathbb{E}_i\left[V_i^1(r)-V_i^0(r)\right] \\
&= \mathbb{E}_i\left[V_i^1(r+\Delta)-V_i^1(r)\right]-\mathbb{E}_i\left[V_i^0(r+\Delta)-V_i^0(r)\right] \\
&= \mathbb{E}_i\int_{x=r}^{x=r+\Delta} q_i^1(x)dx-\mathbb{E}_i\int_{x=r}^{x=r+\Delta} q_i^0(x)dx \\
&= \int_{x=r}^{x=r+\Delta} \mathbb{E}_i q_i^1(x)dx-\int_{x=r}^{x=r+\Delta} \mathbb{E}_i q_i^0(x)dx \\
&= \int_{x=r}^{x=r+\Delta} D(z=1|x)dx.
\end{aligned}$$

This completes the proof of (3). It follows immediately that $\bar{W}'(r)=D(z=1|r)$ at all points of differentiability. The conditions for where $\bar{W}$ is differentiable follow from Lemma A.1.

To prove the statement in (5), note that the assumptions of the Theorem imply that $\bar{\mathscr{V}}$ is differentiable in both r and p by Lemma A.3. In particular, application of Theorem 3 of Milgrom and Segal (2002) in the proof of Lemma A.3 shows that

$$\begin{aligned}
\frac{d}{dr}\bar{\mathscr{V}}(p,r) &= Pr(z=1|p,r) \\
\frac{d}{dp}\bar{\mathscr{V}}(p,r) &= -Pr(j=1|p,r).
\end{aligned}$$

Now, when $Pr(z=1|p,r)$ and $Pr(j=1|p,r)$ are continuously differentiable, the cross-partials $\frac{d}{dp}\frac{d}{dr}\bar{\mathscr{V}}(p,r)$ and $\frac{d}{dr}\frac{d}{dp}\bar{\mathscr{V}}(p,r)$ are continuous and therefore must be equal to each other. This implies that

$$\frac{d}{dp}Pr(z=1|p,r) = -\frac{d}{dr}Pr(j=1|p,r).$$

To prove the last identify in the Theorem, note that

$$\begin{aligned}
Pr(z=1|p,r) &= Pr(j=1|p,r)\left(D(z=1|r)+Pr(z=1|j=0,r)\right)\\
&+Pr(j=0|p,r)Pr(z=1|j=0,r)\\
&= Pr(j=1|p,r)D(z=1|r)+Pr(j=1|p,r)Pr(z=1|j=0,r)\\
&+Pr(j=0|p,r)Pr(z=1|j=0,r)\\
&= Pr(z=1|j=0,r)\left(Pr(j=1|p,r)+Pr(j=0|p,r)\right)\\
&+Pr(j=1|p,r)D(z=1|r)\\
&= Pr(z=1|j=0,r)+Pr(j=1|p,r)D(z=1|r).
\end{aligned}$$

Since $Pr(z=1|j=0,r)$ and $D(z=1|r)$ are not functions of p, we thus have that

$$\frac{d}{dp}Pr(z=1|p,r) = \frac{d}{dp}Pr(j=1|p,r)D(z=1|r).$$

□

A.4 Graphical Illustration

Figure A.1 illustrates the intuition graphically for a representative individual for the case in which the marginal costs are linear. For simplicity, we assume that there are no auxiliary actions, and that $\bar{K}^0(0)=\bar{K}^1(0)=0$. In this case, the likelihood of executing the task equals the chosen level of attention q. In analogy to standard theories of competitive supply, individuals' choice of q with attention technology j is determined by the intersection of the marginal benefit curve r and the marginal cost curve $\frac{\partial}{\partial q}\bar{K}^j$. As in theories of competitive supply, the total surplus of an individual with technology $j=0$ at incentive r is equal to the area of triangle OAD. Similarly, the total surplus of an individual with technology $j=1$ is equal to the area of triangle OAF. Increasing the incentives r by an amount Δ increases surplus by an amount ABCD under technology $j=0$, and by an amount ABEF under technology $j=1$. The change in WTP for technology $j=1$ is thus given by the area DCEF. The area of DCEF is equal to the height, Δ, multiplied by the average of the lengths of DF and

CE, which is

$$\frac{D(z=1|r)+D(z=1|r+\Delta)}{2}.$$

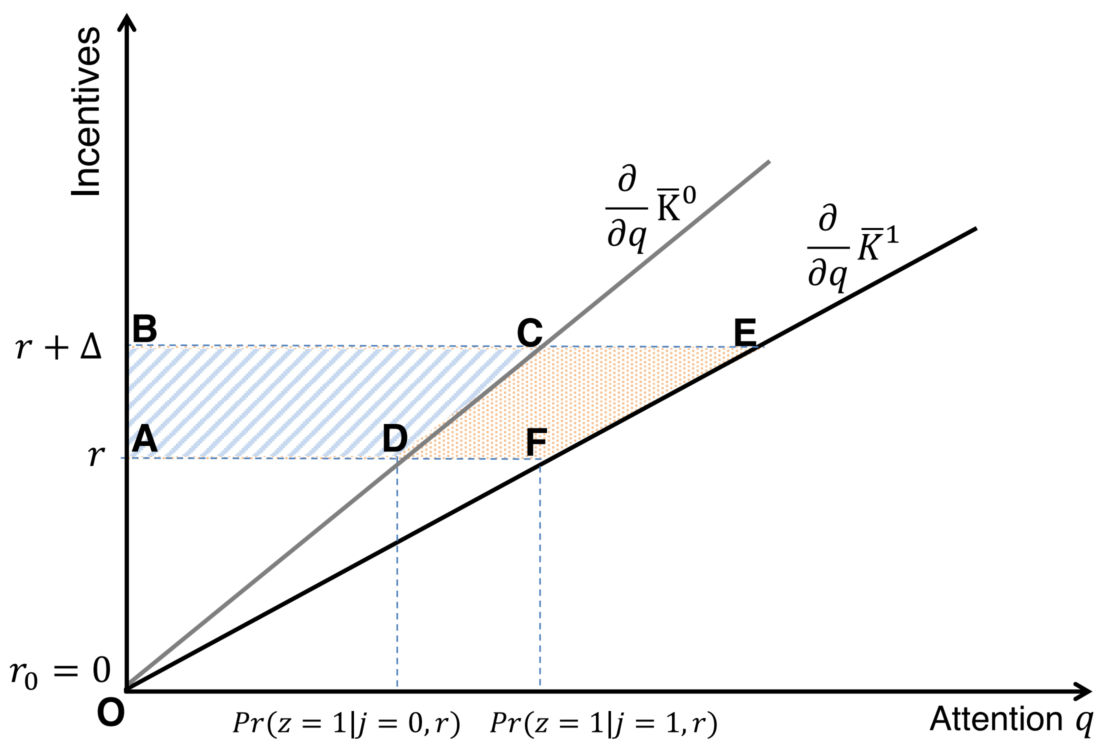
This gives the expression in Corollary 1.

In the limit of very small Δ,

$$\lim_{\Delta\to 0}\frac{\bar{W}(r+\Delta)-\bar{W}(r)}{\Delta}=\lim_{\Delta\to 0}\frac{D(z=1|r)+D(z=1|r+\Delta)}{2}=D(z=1|r)$$

which leads to the first-order condition for $\bar{W}(r)$ in Theorem 1.

Figure A.1: Illustration of Theorem 1



This figure illustrates equation (4) of Theorem 1. The top line (in gray) plots the marginal costs of attention under technology $j=0$, while the bottom line (in black) plots marginal costs under technology $j=1$. The area DCEF corresponds to the change in WTP for technology $j=1$ over $j=0$ when the financial incentive is increased from r to $r+\Delta$.

A.5 Interaction Between Incentives and Reminders

Let $q_{ai}^0(r)$ and $q_{ai}^1(r)$ be the chosen levels of attention given cost functions $\bar{K}_i^0$ and $\bar{K}_i^1$, respectively, and incentive level r. Let $q_{oi}(r)$ denote the auxiliary completion probability conditional on being attentive. Set $\Delta q_{ai}(r) = q_{ai}^1(r) - q_{ai}^0(r)$, and suppose that it is non-negative, meaning that the BE increases attentiveness. The impact of the BE on task completion depends on incentives r as follows:

$$\begin{aligned} \frac{d}{dr} D_i(z=1|r) &= \frac{d}{dr}\left[\Delta q_{ai}(r) \cdot q_{oi}(r)\right] \\ &= \left(\frac{d}{dr}\Delta q_{ai}(r)\right) q_{oi}(r) + \Delta q_{ai}(r) q_{oi}'(r). \end{aligned} \tag{7}$$

Under optimally chosen auxiliary actions, q_{oi} is increasing in r. Moreover, since the BE increases task completion, we have that $\Delta q_{ai}(r) \geq 0$. Thus, equation (7) can be negative only if $\frac{d}{dr}\Delta q_{ai}(r) < 0$, meaning that the BE and incentives are substitutes in people's attention allocation decisions. If attention is chosen optimally, $q_{ai}^j(r)$ is non-decreasing in r, and in fact, any plausible model would make that implication. Combining this property with $\frac{d}{dr}\Delta q_{ai}(r) < 0$ implies that $q_{ai}^0(r)$ must be strictly increasing in r.

B Additional Results for Experiment 1

Table A.1: Participant Characteristics (Experiment 1)

Students		Alumni	
First-year	0.28	2017	0.22
	(0.45)		(0.41)
Sophomore	0.22	2016	0.18
	(0.41)		(0.39)
Junior	0.23	2015	0.21
	(0.42)		(0.41)
Senior	0.27	2014	0.19
	(0.45)		(0.39)
		2013	0.20
			(0.40)
Female	0.65	Female	0.70
	(0.48)		(0.46)
Male	0.32	Male	0.27
	(0.46)		(0.44)
Non-binary or no answer	0.03	Non-binary or no answer	0.03
	(0.18)		(0.18)
N	686	N	687

This table presents summary statistics for the participants in experiment 1, split between student and alumni groups. These participants were randomized to our various treatments as described in the main text. The Pay-to-Code sample includes 496 participants divided between $2 and $5 incentive arms. The Pay-to-Plan sample includes 487 participants divided between $1 and $2 incentive arms. The remaining participants include 218 control participants and 172 participants assigned to the *Combination* treatment. Note that class year is missing for 4 students and 7 alumni, and that gender is missing for 7 students and 9 alumni. The Ns refer to the number of participants in each group. Standard deviations are shown in parentheses.

Table A.2: The Effect of Coding-Task Incentives on Task Completion

	(1) Week 1	(2) Weeks 1-4	(3) Weeks 1-8
>0	0.036	0.032	0.026
	(0.009)	(0.007)	(0.006)
Obs.	714	714	714
R^2	0.039	0.059	0.064
Control Mean	0.385	0.278	0.210
>10	0.037	0.034	0.027
	(0.009)	(0.007)	(0.006)
Obs.	714	714	714
R^2	0.047	0.067	0.072
Control Mean	0.339	0.243	0.179
>30	0.036	0.027	0.023
	(0.009)	(0.006)	(0.005)
Obs.	714	714	714
R^2	0.043	0.053	0.068
Control Mean	0.239	0.186	0.138
>40	0.038	0.026	0.021
	(0.009)	(0.006)	(0.005)
Obs.	714	714	714
R^2	0.044	0.058	0.074
Control Mean	0.183	0.161	0.119
>50	0.032	0.022	0.017
	(0.008)	(0.005)	(0.004)
Obs.	714	714	714
R^2	0.037	0.058	0.071
Control Mean	0.165	0.142	0.107
>60	0.027	0.019	0.013
	(0.008)	(0.005)	(0.004)
Obs.	714	714	714
R^2	0.044	0.066	0.066
Control Mean	0.138	0.118	0.093
Controls	Yes	Yes	Yes
Campus × Student FE	Yes	Yes	Yes

This table presents estimates for the effect of coding-task incentives (in dollars) on task completion. Each panel of the table corresponds to an analysis of whether participants completed at least that number of minutes of the coding task in a given week. The columns correspond to different periods during the experiment over which the effect of the incentives is tested: Column (1) shows the effect in week 1, Column (2) shows the effect for weeks 1-4, and Column (3) shows the effect over all weeks. In Column (1), the dependent variable is an indicator for whether a participant completed at least that many minutes of the coding task in the first week. In Columns (2) and (3), the dependent variable is the mean of the indicators, constructed as in Column (1), for each of the weeks being considered. Each panel-by-column corresponds to a separate specification, and thus 18 distinct specifications are shown in the table. Standard errors, clustered at the participant level, are shown in parentheses.

Table A.3: The Effect of Plan-Making Incentives on Task Completion

	(1) Week 1	(2) Weeks 1-4	(3) Weeks 1-8
>0	0.037	0.029	0.014
	(0.022)	(0.015)	(0.012)
Obs.	705	705	705
R^2	0.041	0.040	0.046
Control Mean	0.385	0.278	0.210
>10	0.037	0.027	0.014
	(0.021)	(0.014)	(0.011)
Obs.	705	705	705
R^2	0.045	0.045	0.046
Control Mean	0.339	0.243	0.179
>30	0.045	0.023	0.010
	(0.020)	(0.013)	(0.010)
Obs.	705	705	705
R^2	0.054	0.042	0.045
Control Mean	0.239	0.186	0.138
>40	0.036	0.019	0.008
	(0.018)	(0.012)	(0.009)
Obs.	705	705	705
R^2	0.034	0.036	0.041
Control Mean	0.183	0.161	0.119
>50	0.034	0.015	0.005
	(0.018)	(0.011)	(0.008)
Obs.	705	705	705
R^2	0.035	0.039	0.042
Control Mean	0.165	0.142	0.107
>60	0.027	0.013	0.002
	(0.016)	(0.010)	(0.008)
Obs.	705	705	705
R^2	0.044	0.038	0.042
Control Mean	0.138	0.118	0.093
Controls	Yes	Yes	Yes
Campus × Student FE	Yes	Yes	Yes

This table presents estimates for the effect of plan-making incentives (in dollars) on task completion. Each panel of the table corresponds to an analysis of whether participants completed at least that number of minutes of the task in a given week. The columns correspond to different periods during the experiment over which the effect of the incentives is tested: Column (1) shows the effect in week 1, Column (2) shows the effect for weeks 1-4, and Column (3) shows the effect for all weeks. In Column (1), the dependent variable is an indicator for whether a participant completed at least that many minutes of the coding task in the first week. In Columns (2) and (3), the dependent variable is the mean of the indicators, constructed as in Column (1), for each of the weeks being considered. Each panel-by-column corresponds to a separate specification, and thus 18 distinct specifications are shown in the table. Standard errors, clustered at the participant level, are shown in parentheses.

Table A.4: The Effect of Plan-Making Incentives on Task Completion (2SLS)

A. The Effect on Plan Making (First Stage)

	(1) Week 1	(2) Weeks 1-4	(3) Weeks 1-8
$1 Plan	0.282	0.285	0.240
	(0.048)	(0.033)	(0.030)
$2 Plan	0.368	0.297	0.242
	(0.033)	(0.028)	(0.025)
Obs.	705	705	705
R^2	0.144	0.189	0.157
Control Mean	0.381	0.150	0.082
Controls	Yes	Yes	Yes
Campus FE	Yes	Yes	Yes

B. The Effect on Coding Task Completion (Reduced Form)

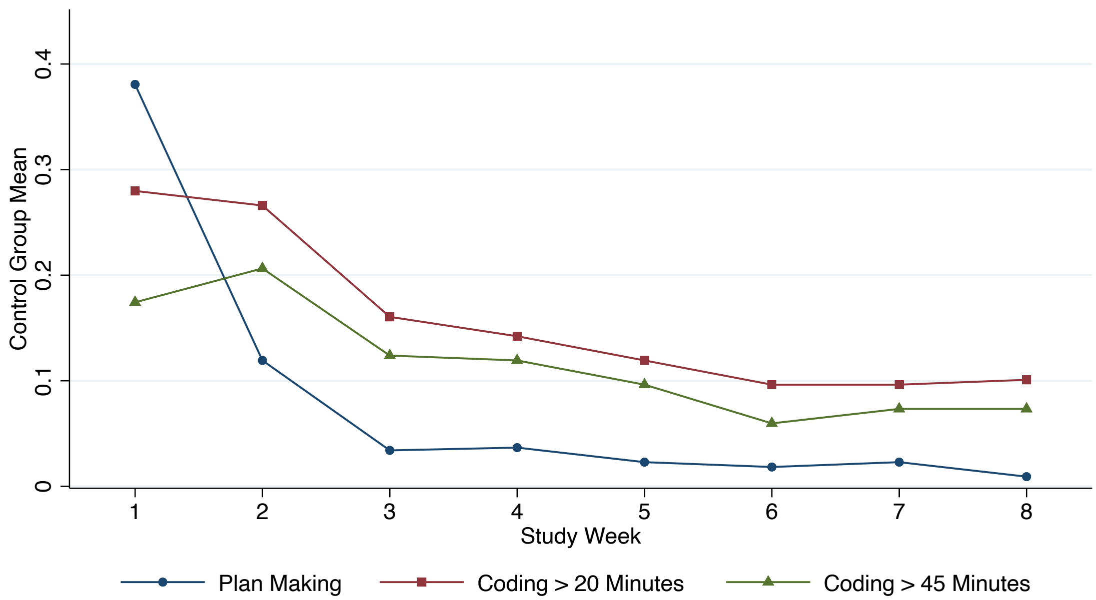
	(1) >20 (1)	(2) >20 (1-4)	(3) >20 (1-8)	(4) >45 (1)	(5) >45 (1-4)	(6) >45 (1-8)
$1 Plan	0.034	0.017	0.013	0.034	-0.000	0.005
	(0.048)	(0.032)	(0.025)	(0.043)	(0.027)	(0.021)
$2 Plan	0.079	0.054	0.026	0.076	0.032	0.012
	(0.040)	(0.027)	(0.021)	(0.036)	(0.023)	(0.018)
Obs.	705	705	705	705	705	705
R^2	0.057	0.049	0.051	0.036	0.035	0.041
Control Mean	0.280	0.212	0.158	0.174	0.156	0.116
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Campus FE	Yes	Yes	Yes	Yes	Yes	Yes

C. The Effect of Plan Making on Coding Task Completion (IV)

	(1) >20 (1)	(2) >20 (1-4)	(3) >20 (1-8)	(4) >45 (1)	(5) >45 (1-4)	(6) >45 (1-8)
Plan Making	0.203	0.146	0.092	0.197	0.076	0.041
	(0.102)	(0.080)	(0.078)	(0.093)	(0.070)	(0.066)
Obs.	705	705	705	705	705	705
R^2	0.143	0.151	0.120	0.091	0.094	0.076
Control Mean	0.280	0.212	0.158	0.174	0.156	0.116
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Campus FE	Yes	Yes	Yes	Yes	Yes	Yes

This table shows estimates for the effect of plan-making incentives on plan making and task completion using treatment dummies rather than a linear plan-making incentive variable. Panel A shows estimates of the effect of plan-making incentives on whether or not participants made a plan. Column (1) shows the effect of plan-making incentives in week 1. Column (2) shows the average effect over weeks 1-4. Column (3) shows the average effect over all weeks. Panel B shows the effect of plan-making incentives on task completion. Columns (1)–(3) show the effect on an indicator variable for whether or not the participant worked on the coding task for more than 20 minutes: Column (1) estimates the effect over week 1, Column (2) over weeks 1-4, and Column (3) over all weeks. Columns (4)–(6) show analogous estimates, but for an indicator variable for whether or not the participant worked on the task for more than 45 minutes each week. Panel C shows the 2SLS estimates instrumenting for whether or not participants made a plan using the plan-making treatment dummies as instruments. The dependent variables are the same as those in Panel B. Standard errors, clustered at the participant level, are shown in parentheses.

Figure A.2: Experiment 1 Control Group Means (Week-by-Week)



This figure shows control group means for plan making and completing at least 20 minutes or at least 45 minutes of the coding course for each week of the study.

C Additional Results for Experiment 2

Table A.5: The Effect of Incentive, Delay, and Reminders on Part-2 Survey Completion (Categorical Delay)

	Completed Part-2 Survey		
	(1)	(2)	(3)
Received Reminder	0.23	0.15	0.16
	(0.021)	(0.045)	(0.044)
High Incentive	0.08	0.14	0.13
	(0.021)	(0.029)	(0.031)
1-Week Delay	-0.16	-0.22	-0.22
	(0.029)	(0.041)	(0.072)
3-Week Delay	-0.17	-0.27	-0.27
	(0.030)	(0.042)	(0.067)
6-Week Delay	-0.21	-0.32	-0.30
	(0.030)	(0.041)	(0.063)
Received Reminder × High Incentive		-0.13	-0.11
		(0.042)	(0.040)
1-Week Delay × Received Reminder		0.13	0.10
		(0.058)	(0.075)
3-Week Delay × Received Reminder		0.20	0.20
		(0.060)	(0.057)
6-Week Delay × Received Reminder		0.22	0.18
		(0.059)	(0.063)
Constant	0.55	0.59	0.59
	(0.025)	(0.032)	(0.043)
Observations	2,076	2,076	2,076
Number of Participants	2,076	2,076	2,076
S.E. Clustered by P1 & P2 Date			X
P1 Date FE			X

This table estimates how survey completion varies with reminders, delay, and whether participants are offered high incentives (i.e., \$11 or \$12) or low incentives (i.e., \$3 or \$4) to complete the survey. The 2-day delay variable is omitted so the 2-day delay is the excluded group. This table only includes the 90% of participants who were randomly assigned to receive or not receive reminders. Column (3) reproduces Column (2) with fixed effects for the date that part 1 of the study was taken and with standard errors clustered for the date the participant completed part 1 and the date part 2 was made available to them. Standard errors are shown in parentheses.

Table A.6: The Effect of Incentive and Delay on Willingness to Pay for Reminders (Only Participants Randomized for Reminders)

	WTP for Reminders ($)					
	(1)	(2)	(3)	(4)	(5)	(6)
Extra $1	0.07	0.08	0.06	0.07	0.10	0.11
	(0.018)	(0.055)	(0.038)	(0.018)	(0.019)	(0.055)
High Incentive	0.96	0.96	1.14	0.96	0.96	0.96
	(0.082)	(0.082)	(0.118)	(0.082)	(0.082)	(0.082)
Extra $1 × High Incentive	-0.05	-0.05	-0.01	-0.05	-0.06	-0.06
	(0.050)	(0.050)	(0.116)	(0.050)	(0.051)	(0.051)
Ln(P2 Delay)		-0.07	-0.04			-0.07
		(0.022)	(0.013)			(0.023)
Extra $1 × Ln(P2 Delay)		-0.00	0.01			-0.00
		(0.022)	(0.014)			(0.022)
High Incentive × Ln(P2 Delay)			-0.08			
			(0.036)			
Extra $1 × Ln(P2 Delay)			-0.02			
× High Incentive			(0.044)			
Constant	0.51	0.68	0.59	0.51	0.51	0.68
	(0.034)	(0.060)	(0.044)	(0.034)	(0.034)	(0.060)
Observations	33,216	33,216	33,216	33,216	33,216	33,216
Number of Participants	2,076	2,076	2,076	2,076	2,076	2,076
Specification	OLS	OLS	OLS	OLS	Tobit	Tobit
P1 Date FE				X		

This table estimates how willingness to pay for reminders varies with the natural log of delay (in days) and incentives to complete the survey. This table only includes participants who were randomly assigned to receive or not receive reminders. The High Incentive variable is an indicator for being asked about an incentive of $11 or $12. The Extra $1 variable is an indicator for being asked about an incentive of $4 or $12. Column (4) reproduces Column (1) with fixed effects for the date that part 1 of the survey was taken; Columns (5) and (6) reproduce Columns (1) and (2) using Tobit estimates with censors at -$4 and $4 for the low-incentive group and censors at -$12 and $12 for the high-incentive group. Standard errors, clustered at the participant level, are shown in parentheses.

Table A.7: The Effect of Incentive and Delay on Willingness to Pay for Reminders by MPL

	WTP for Reminders ($)			
	(1)	(2)	(3)	(4)
Extra $1	0.15	0.08	0.11	0.09
	(0.078)	(0.050)	(0.034)	(0.026)
High Incentive	0.58	0.68	0.79	0.77
	(0.193)	(0.154)	(0.143)	(0.139)
Extra $1 $\times$ High Incentive	0.18	0.02	-0.10	-0.02
	(0.244)	(0.150)	(0.106)	(0.077)
Constant	0.45	0.49	0.50	0.51
	(0.063)	(0.053)	(0.048)	(0.046)
Observations	4,612	9,224	13,836	18,448
Number of Participants	2,306	2,306	2,306	2,306
First T MPLs	$T=2$	$T=4$	$T=6$	$T=8$

This table estimates how willingness to pay for reminders varies with incentives to complete the survey. The High Incentive variable is an indicator for being asked about an incentive of $11 or $12. The Extra $1 variable is an indicator for being asked about an incentive of $4 or $12. Column (1) shows these estimates when limited to the first 2 MPLs participants are asked about, Column (2) shows these estimates when limited to the first 4 MPLs, Column (3) shows these estimates when limited to the first 6 MPLs, and Column (4) shows these estimates when limited to the first 8 MPLs. Standard errors, clustered at the participant level, are shown in parentheses.

Table A.8: The Effect of Incentive and Delay on WTP for Reminders (Categorical Delay)

	WTP for Reminders ($)				
	(1)	(2)	(3)	(4)	(5)
Extra $1	0.07	0.05	0.07	0.10	0.08
	(0.017)	(0.047)	(0.017)	(0.018)	(0.047)
High Incentive	0.93	0.93	0.93	0.93	0.93
	(0.077)	(0.077)	(0.077)	(0.077)	(0.077)
Extra $1 × High Incentive	-0.06	-0.06	-0.06	-0.07	-0.07
	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)
1-Week Delay		-0.16			-0.16
		(0.054)			(0.055)
3-Week Delay		-0.19			-0.19
		(0.062)			(0.062)
6-Week Delay		-0.25			-0.25
		(0.067)			(0.067)
1-Week Delay × Extra $1		0.03			0.03
		(0.073)			(0.074)
3-Week Delay × Extra $1		0.02			0.02
		(0.068)			(0.068)
6-Week Delay × Extra $1		0.03			0.03
		(0.068)			(0.068)
Constant	0.51	0.66	0.51	0.50	0.65
	(0.032)	(0.047)	(0.032)	(0.032)	(0.047)
Observations	36,896	36,896	36,896	36,896	36,896
Number of Participants	2,306	2,306	2,306	2,306	2,306
Specification	OLS	OLS	OLS	Tobit	Tobit
P1 Date FE			X		

This table estimates how willingness to pay for reminders varies with incentives to complete the survey. The High Incentive variable is an indicator for being asked about an incentive of $11 or $12. The Extra $1 variable is an indicator for being asked about an incentive of $4 or $12. Column (2) maintains the specification in Column (1) and adds controls for delay; Column (3) shows Column (1) with fixed effects for the date that part 1 of the survey was taken; Columns (4) and (5) reproduce Columns (1) and (2) using Tobit estimates with censors at -$4 and $4 for the low-incentive group and censors at -$12 and $12 for the high-incentive group. Standard errors, clustered at the participant level, are shown in parentheses.

D Additional Results for Experiment 3

Table A.9: Replication of Table 6 with Tobit Models

	Willingness to Pay ($)					
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive ($)	0.11	-0.00	0.05	0.03	0.08	0.03
	(0.035)	(0.031)	(0.046)	(0.042)	(0.052)	(0.073)
Incentive ($) × Block 2	-0.07	0.04	0.01	0.04	0.00	0.02
	(0.038)	(0.037)	(0.051)	(0.052)	(0.065)	(0.076)
Incentive ($) × Feedback			0.11	-0.07	0.12	0.11
			(0.069)	(0.061)	(0.093)	(0.100)
Incentive ($) × Block 2			-0.16	0.01	-0.22	-0.11
× Feedback			(0.076)	(0.075)	(0.103)	(0.110)
Block 2	0.14	-0.18	-0.10	-0.21	-0.17	-0.06
	(0.126)	(0.124)	(0.178)	(0.177)	(0.236)	(0.260)
Block 2 × Feedback			0.49	0.05	0.89	0.17
			(0.251)	(0.248)	(0.343)	(0.359)
Feedback			-0.23	0.22	-0.37	-0.12
			(0.250)	(0.214)	(0.347)	(0.354)
Constant	0.17	0.51	0.29	0.40	0.28	0.30
	(0.125)	(0.107)	(0.172)	(0.148)	(0.212)	(0.260)
Observations	3,996	4,794	3,996	4,794	1,788	2,208
Number of Participants	666	799	666	799	298	368
Participant B1 Acc. Diff.	All	All	All	All	≤ 0	> 0
Arm	Length	Discernibility	Length	Discernibility	Length	Length

This table replicates Table 6 but presents Tobit estimates with censors at $4 and -$4 for the effect of accuracy incentive, block order, and whether the participant received feedback on their block-1 performance on willingness to pay for an easy task (i.e., a task with shorter length in the length arm, or a task with increased discernibility in the discernibility arm). Standard errors, clustered at the participant level, are shown in parentheses.

Table A.10: Replication of Table 6, Dropping the 10% Fastest Participants

	Willingness to Pay ($)					
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive ($)	0.10	-0.01	0.04	0.02	0.08	0.01
	(0.037)	(0.033)	(0.051)	(0.046)	(0.061)	(0.076)
Incentive ($) × Block 2	-0.07	0.03	0.02	0.03	0.03	0.02
	(0.041)	(0.040)	(0.055)	(0.056)	(0.075)	(0.078)
Incentive ($) × Feedback			0.12	-0.07	0.14	0.12
			(0.073)	(0.066)	(0.096)	(0.103)
Incentive ($) × Block 2			-0.19	0.00	-0.31	-0.12
× Feedback			(0.082)	(0.081)	(0.117)	(0.112)
Block 2	0.16	-0.12	-0.13	-0.15	-0.24	-0.06
	(0.138)	(0.133)	(0.198)	(0.189)	(0.277)	(0.275)
Block 2 × Feedback			0.59	0.08	1.24	0.20
			(0.274)	(0.266)	(0.397)	(0.371)
Feedback			-0.15	0.17	-0.30	-0.06
			(0.270)	(0.231)	(0.389)	(0.369)
Constant	0.17	0.50	0.24	0.42	0.18	0.28
	(0.135)	(0.115)	(0.192)	(0.160)	(0.252)	(0.275)
Observations	3,438	4,314	3,438	4,314	1,344	2,094
Number of Participants	573	719	573	719	224	349
Participant B1 Acc. Diff.	All	All	All	All	≤ 0	> 0
Arm	Length	Discernibility	Length	Discernibility	Length	Length

This table replicates Table 6 on the effect of accuracy incentive, block order, and whether the participant received feedback on their block-1 performance on willingness to pay for an easy task (i.e., a task with shorter length in the length arm, or a task with increased discernibility in the discernibility arm) after dropping participants in the top 10% of fastest task times in the length arm and the top 10% of fastest task times in the discernibility arm. Standard errors, clustered at the participant level, are shown in parentheses.

Table A.11: The Effect of Block and Feedback on Block Accuracy Difference

	Accuracy Difference Between Easy and Baseline Tasks					
	(1)	(2)	(3)	(4)	(5)	(6)
Block 2	0.03	0.01	0.20	-0.13	0.19	-0.10
	(0.021)	(0.018)	(0.018)	(0.020)	(0.027)	(0.029)
Feedback	0.03	-0.01			0.02	0.03
	(0.023)	(0.020)			(0.016)	(0.023)
Block 2 × Feedback	-0.04	-0.01			0.00	-0.07
	(0.030)	(0.025)			(0.036)	(0.039)
Constant	0.18	0.26	-0.06	0.40	-0.07	0.39
	(0.016)	(0.014)	(0.008)	(0.012)	(0.013)	(0.016)
Observations	1,332	1,598	596	736	596	736
Number of Participants	666	799	298	368	298	368
Participant B1 Acc. Diff.	All	All	≤ 0	> 0	≤ 0	> 0
Arm	Length	Discernibility	Length	Length	Length	Length

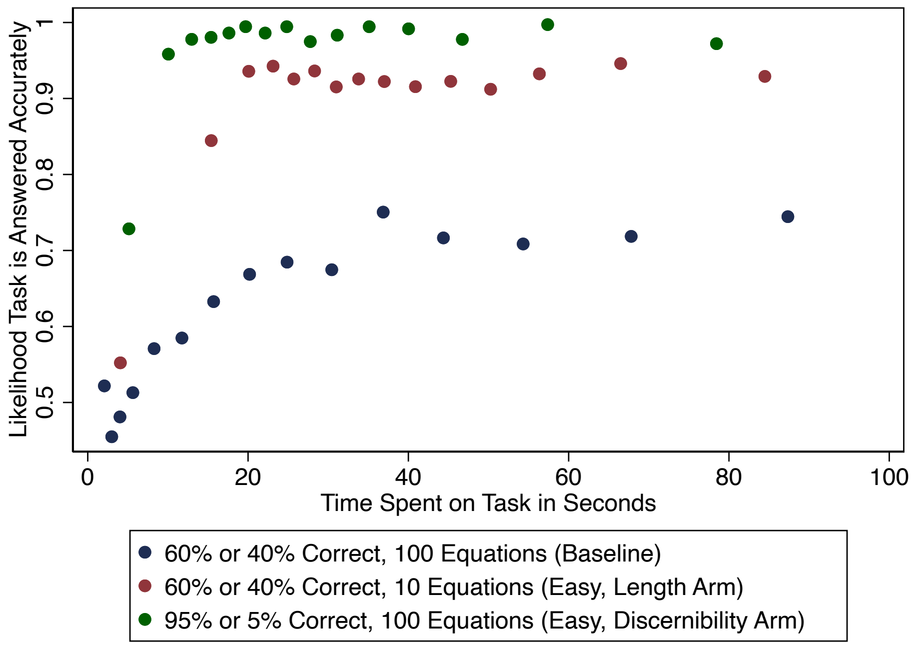
This table estimates the effect of block order and feedback on the accuracy difference between easier and baseline tasks within a block. The accuracy difference is constructed by taking the difference between the percentage of easy tasks answered correctly and the percentage of baseline tasks answered correctly in a block. Column (1) shows OLS estimates for participants in the length arm; Column (2) shows OLS estimates for participants in the discernibility arm; Columns (3) and (5) restrict to participants in the length arm who had a block-1 accuracy difference less than or equal to 0 (i.e., who were at least as accurate in the baseline tasks as in the easier tasks); Columns (4) and (6) restrict to participants in the length arm who had a block-1 accuracy difference greater than 0 (i.e., who were more accurate in the easier tasks than the baseline tasks). Standard errors, clustered at the participant level, are shown in parentheses.

Table A.12: The Effect of Block, Feedback, and Accuracy Difference on Within-Block Time Spent on Baseline vs. Easy Tasks (Length Arm)

	Average Difference in Time Spent on Baseline vs. Easy Tasks, By Block					
	(1)	(2)	(3)	(4)	(5)	(6)
Block 2	-21.72	-13.78	-19.24	-11.59	-16.63	-9.04
	(4.125)	(6.078)	(3.562)	(5.444)	(3.144)	(4.740)
B1 Acc. Diff ≤ 0	115.45	106.94	103.16	93.86	79.81	74.50
	(10.006)	(14.582)	(8.194)	(11.926)	(6.141)	(9.036)
Block 2 $\times$ B1 Acc. Diff ≤ 0	-43.62	-54.99	-33.48	-42.74	-16.53	-26.72
	(7.547)	(11.468)	(5.928)	(8.959)	(4.665)	(6.987)
Feedback		3.53		3.40		6.12
		(9.958)		(8.729)		(7.343)
Block 2 $\times$ Feedback		-16.07		-15.47		-15.34
		(8.211)		(7.074)		(6.236)
B1 Acc. Diff ≤ 0 $\times$ Feedback		17.26		18.86		10.78
		(19.988)		(16.347)		(12.243)
Block 2 $\times$ B1 Acc. Diff ≤ 0 $\times$ Feedback		23.01		18.74		20.61
		(15.056)		(11.823)		(9.290)
Constant	19.23	17.48	16.51	14.83	11.79	8.76
	(4.978)	(7.384)	(4.366)	(6.615)	(3.673)	(5.423)
Observations	1,332	1,332	1,332	1,332	1,332	1,332
Number of Participants	666	666	666	666	666	666
Winsorized at T Seconds	No	No	$T = 300$	$T = 300$	$T = 180$	$T = 180$

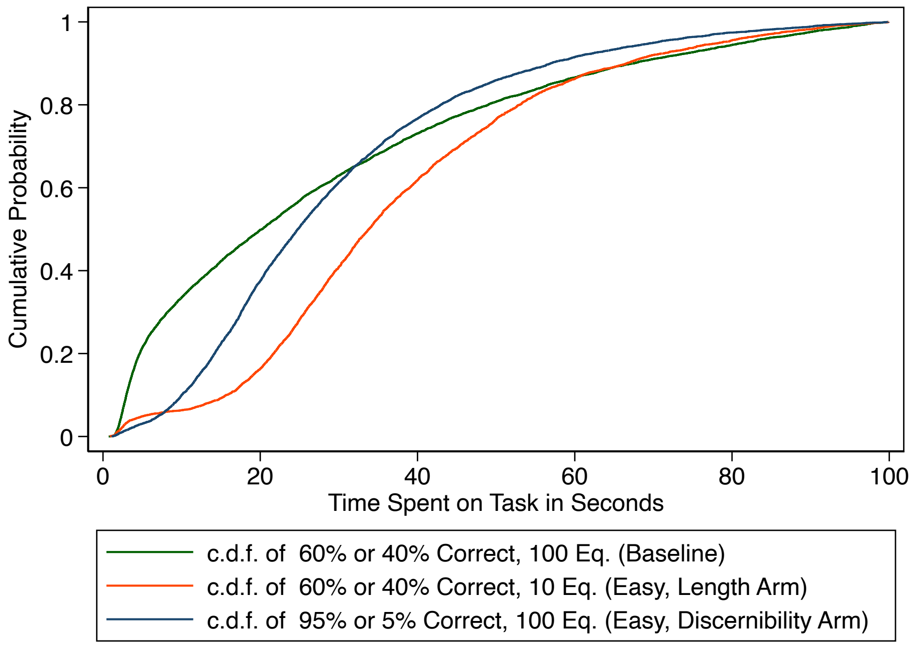
This table estimates the effect of block, feedback and accuracy difference on the average difference in time spent on baseline vs. easy tasks for participants in the length arm. The dependent variable is constructed by taking the difference between average time spent on the baseline tasks in a block and the average time spent on the easy tasks in the same block. By-block accuracy difference is constructed by taking the difference between the percentage of easy tasks answered correctly and the percentage of baseline tasks answered correctly in a block. Column (3) maintains the specification in Column (1) and winsorizes at 300 seconds; Column (4) maintains the specification in Column (2) and winsorizes at 300 seconds; Column (5) maintains the specification in Column (1) and winsorizes at 180 seconds; Column (6) maintains the specification in Column (2) and winsorizes at 180 seconds. Standard errors, clustered at the participant level, are shown in parentheses.

Figure A.3: The Likelihood of Answering Tasks Accurately by Time Spent, Dropping Observations > 100 Seconds



This figure includes a binned scatterplot that displays how accuracy varies with the time spent on the three types of tasks in seconds after dropping responses in which a participant spent more than 100 seconds on a task. Here, observations have been separated into 15 equal-sized bins.

Figure A.4: CDF of Time Spent on Task by Length and Discernibility (Dropping Observations > 100 Seconds)



This figure displays the CDF of time spent on a task in seconds by task type after dropping responses in which a participant spent more than 100 seconds on a task.

E Implications of Risk Aversion

E.1 Calibration exercises

Let π_1 denote the probability of completing the task with BE, and let π_0 denote the probability of completing the task without the BE. Suppose that individuals value the BE optimally but are potentially risk averse. Let z denote the initial wealth of an individual before starting the experiment. We consider two cases.

Case 1: Constant absolute risk aversion Let α be the CARA parameter. Analogous to before, an application of the Envelope Theorem implies that a marginal increase dr in r increases the individual's expected utility by $\pi_1 e^{-\alpha(z+r)}dr$ in the presence of the BE, and by $\pi_0 e^{-\alpha(z+r)}dr$ in the absence of a BE. A marginal increase dw in w, where $w > 0$ is a payment received in the absence of the BE, increases an individual's expected utility by $\pi_0 e^{-\alpha(z+w+r)}dw + (1-\pi_0)e^{-\alpha(z+w)}dw$ in the absence of the BE. Dividing through by $e^{-\alpha z}$

implies that the the impact of r on the WTP for a BE is given by

$$\frac{dw}{dr} = \frac{\pi_1 e^{-\alpha r} - \pi_0 e^{-\alpha r}}{\pi_0 e^{-\alpha(w+r)} + (1-\pi_0)e^{-\alpha w}}. \tag{8}$$

We use equation (8) to estimate how WTP for a BE changes when the reward for (accurate) task completion increases by \$1. In experiment 2, we set $r = \$3$ for the low incentive conditions and $r = \$11$ for the high incentive conditions. We set π_0 and π_1 to match the empirical completion probabilities in the 8 delay×incentive conditions in Figure 2a. We set w to match the average WTP for the BE at the \$3 or \$11 incentive values in the 8 delay×incentive conditions in Figure 3.

In experiment 3, we set $r = \$2$, and we set w to match the average WTP for the BE at the $r = \$2$ incentive value, in each of the 4 conditions corresponding to either the length or discernibility arm, and either block 1 or block 2.

We draw prior work to consider the following values of α: Using insurance decisions, Cohen and Einav (2007) estimate $\alpha \approx (0.00087, 0.0019)$, Handel (2013) estimates $\alpha \approx (0.00019, 0.000325)$, and Sydnor (2010) estimates $\alpha \approx 0.002$. Chetty (2006) estimates a constant relative risk aversion coefficient of 0.7 from labor supply elasticities, which translates to $\alpha \approx 0.0007$ if payday borrowers have \$1000 monthly "uncommitted" (in the sense of Chetty and Szeidl, 2007) consumption. Using relatively small-stakes gambles, von Gaudecker et al. (2011) estimate $\alpha \approx 0.03$, and Holt and Laury (2002) estimate $\alpha \approx 0.2$. For studies that provide a range, we take the midpoint of the range.

Table A.13 below considers experiment 2 data under the hypothesis that people value BEs optimally but are risk averse. The table presents estimates of how the WTP for the BE would change when r increases by \$1, across the 8 delay×incentive conditions, and across the different values of α summarized above. We set $\alpha = 0$ in the first row, to benchmark to the quasilinear case assumed in the body of the paper. The very last row in the table, separated by double lines, presents our empirical estimates.

Table A.13: Experiment 2, Effect of Extra $1 Incentive on WTP by Delay × Incentive Condition and α (CARA parameter)

	2 Day × Low Incentive	1 Week × Low Incentive	3 Weeks × Low Incentive	6 Weeks × Low Incentive	2 Day × High Incentive	1 Week × High Incentive	3 Weeks × High Incentive	6 Weeks × High Incentive
$\alpha = 0$	0.23	0.28	0.40	0.25	−0.07	0.16	0.18	0.36
$\alpha \approx (0.00087, 0.0019)$ Cohen and Einav (2007)	(0.23, 0.23)	(0.28, 0.28)	(0.40, 0.40)	(0.25, 0.24)	(−0.07, −0.07)	(0.16, 0.16)	(0.18, 0.18)	(0.36, 0.36)
$\alpha \approx (0.00019, 0.000325)$ Handel (2013)	(0.23, 0.23)	(0.28, 0.28)	(0.40, 0.40)	(0.25, 0.25)	(−0.07, −0.07)	(0.16, 0.16)	(0.18, 0.18)	(0.36, 0.36)
$\alpha \approx 0.002$ Sydnor (2010)	0.23	0.28	0.40	0.24	−0.07	0.16	0.18	0.36
$\alpha \approx 0.0007$ Chetty (2006)	0.23	0.28	0.40	0.25	−0.07	0.16	0.18	0.36
$\alpha \approx 0.03$ von Gaudecker et al. (2011)	0.23	0.27	0.38	0.23	−0.07	0.14	0.15	0.30
$\alpha \approx 0.2$ Holt and Laury (2002)	0.19	0.21	0.27	0.17	−0.03	0.04	0.05	0.07
Empirical estimates	0.06	0.10	0.07	0.08	−0.00	0.01	0.03	0.03

Notes: This table presents estimates of how the WTP for the BE would change when r increases by $1, across the 8 delay × incentive conditions and across the values of α summarized above.

Table A.14 below considers experiment 3 data under the hypothesis that people value BEs optimally but are risk averse. The table presents estimates of how the WTP for the BE changes when r increases by $1, across the 4 conditions corresponding to either the length or discernibility arm, and either block 1 or block 2. As in Table A.13, we consider the different values of α summarized above, as well as the risk-neutral $\alpha = 0$ in the first row. The very last row in the table, separated by double lines, presents our empirical estimates.

Table A.14: Experiment 3, Effect of Extra \$1 Incentive on WTP by Arm $\times$ Block and α (CARA parameter)

	Length Arm $\times$ Block 1	Discernibility Arm $\times$ Block 1	Length Arm $\times$ Block 2	Discernibility Arm $\times$ Block 2
$\alpha = 0$	0.20	0.26	0.21	0.26
$\alpha \approx (0.00087, 0.0019)$ Cohen and Einav (2007)	(0.20, 0.20)	(0.26, 0.26)	(0.21, 0.21)	(0.26, 0.26)
$\alpha \approx (0.00019, 0.000325)$ Handel (2013)	(0.20, 0.20)	(0.26, 0.26)	(0.21, 0.21)	(0.26, 0.26)
$\alpha \approx 0.002$ Sydnor (2010)	0.20	0.26	0.21	0.26
$\alpha \approx 0.0007$ Chetty (2006)	0.20	0.26	0.21	0.26
$\alpha \approx 0.03$ von Gaudecker et al. (2011)	0.19	0.26	0.21	0.26
$\alpha \approx 0.2$ Holt and Laury (2002)	0.19	0.25	0.20	0.25
Empirical estimates	0.10	−0.01	0.03	0.03

Notes: This table presents estimates of how the WTP for the BE would change when r increases by \$1, across the 4 arm × block conditions and across the values of α summarized above.

Case 2: Constant relative risk aversion Let ρ be the CRRA parameter. Analogous to above, simple algebra shows that

$$\frac{dw}{dr} = \frac{\pi_1 - \pi_0}{\pi_0 \left(\frac{z+r}{z+w+r}\right)^{\rho} + (1-\pi_0)\left(\frac{z+r}{z+w}\right)^{\rho}}. \tag{9}$$

To study the potential impacts of risk aversion, we consider the upper-bound value of $\rho = 1.37$, which Holt and Laury (2002) clarify implies a level of risk aversion that individuals with such a parameter should "stay in bed." Holt and Laury (2002) show that very few individuals exhibit such a value of risk aversion.

Table A.15 below considers experiment 2 data under the hypothesis that people value BEs optimally but are risk averse. Utilizing equation (9), table presents estimates of how the WTP for the BE would change when r increases by \$1, across the 8 delay$\times$incentive conditions.

The first row corresponds to the risk-neutral benchmark of $\rho = 0$. The subsequent rows consider $\rho = 1.37$ and vary assumptions about initial wealth z. The very last row in the table, separated by double lines, presents our empirical estimates.

Table A.16 is analogous, but considers the 4 conditions corresponding to either the length or discernibility arm, and either block 1 or block 2.

Table A.15: Experiment 2, Effect of Extra $1 Incentive on WTP by Delay × Incentive Condition, initial wealth and ρ (CRRA parameter)

	2 Day × Low Incentive	1 Week × Low Incentive	3 Weeks × Low Incentive	6 Weeks × Low Incentive	2 Day × High Incentive	1 Week × High Incentive	3 Weeks × High Incentive	6 Weeks × High Incentive
$\rho = 0$	0.23	0.28	0.40	0.25	−0.07	0.16	0.18	0.36
$\rho = 1.37, z = 10$	0.21	0.24	0.32	0.20	−0.06	0.10	0.11	0.19
$\rho = 1.37, z = 100$	0.23	0.28	0.39	0.24	−0.07	0.15	0.17	0.34
$\rho = 1.37, z = 1000$	0.23	0.28	0.40	0.25	−0.07	0.16	0.18	0.36
$\rho = 1.37, z = 100000$	0.23	0.28	0.40	0.25	−0.07	0.16	0.18	0.36
Empirical estimates	0.06	0.10	0.07	0.08	−0.00	0.01	0.03	0.03

Notes: This table presents estimates of how the WTP for the BE would change when r increases by $1, across the 8 delay × incentive conditions and across the values of ρ and z summarized above.

Table A.16: Experiment 3, Effect of Extra $1 Incentive on WTP by Delay $\times$ Incentive Condition, initial wealth and ρ (CRRA parameter)

	Length Arm $\times$ Block 1	Discernibility Arm $\times$ Block 1	Length Arm $\times$ Block 2	Discernibility Arm $\times$ Block 2
$\rho = 0$	0.20	0.26	0.21	0.26
$\rho = 1.37, z = 10$	0.19	0.25	0.20	0.25
$\rho = 1.37, z = 100$	0.19	0.26	0.21	0.26
$\rho = 1.37, z = 1000$	0.20	0.26	0.21	0.26
$\rho = 1.37, z = 100000$	0.20	0.26	0.21	0.26
Empirical estimates	0.10	−0.01	0.03	0.03

Notes: This table presents estimates of how the WTP for the BE would change when r increases by $1, across the 4 arm × block conditions and across the values of ρ and z summarized above.

E.2 Risk Aversion and Willingness to Pay in Experiment 2

Appendix Table A.17 explores whether the deviations in experiment 2 between willingness to pay and the effect of reminders on task completion depend on participant risk aversion. People value reminders by selecting between a sure amount of money and a reminder that only increases the likelihood of a monetary reward. If risk aversion is sufficiently high, then risk aversion, rather than suboptimal attention, could account for these results.

We estimate how willingness to pay for reminders varies with the participant's level of risk aversion as measured in part 2 of experiment 2 and with completion of the survey. The risk aversion variables are derived from participant answers to 10 questions in part 2 of experiment 2, in which participants were asked to select between receiving a "for sure" amount of money and receiving a higher amount of money with 50% probability (see Figure A.5 for an example). The Above Median Risky Choice Total variable is an indicator that takes the value of 1 if the number of times that a participant selected the uncertain choice is higher than the sample median. The Fraction of Risky Choices variable is the number of times a participant selected the uncertain choice divided by the total number of questions.

Column (1) shows that there is no difference in willingness to pay for reminders when comparing participants with relatively high versus low risk aversion in terms of the number of risky choices they select in part 2. These results hold both for the low incentive and high incentive groups. Column (2) recasts these results using the fraction of risky choices selected as the interaction term. Again, we see no interaction effect, which suggests that risk aversion is not driving participants' willingness to pay for reminders. In the case of the low incentive group, we can make an even stronger statement: an extreme risk-seeking individual (i.e., with the fraction of risky choices equal to 1) would still value reminders with a 95% confidence interval well below the estimated effect of reminders on task completion.[1] Finally, column (3) confirms that the subset of participants who complete part 2, and thus for whom we can measure risk aversion, have statistically indistinguishable willingness to pay for reminders as compared to the participants who do not complete part 2. This fact supports our extrapolation of the risk aversion results from columns (1) and (2) to characterize the full participant pool.

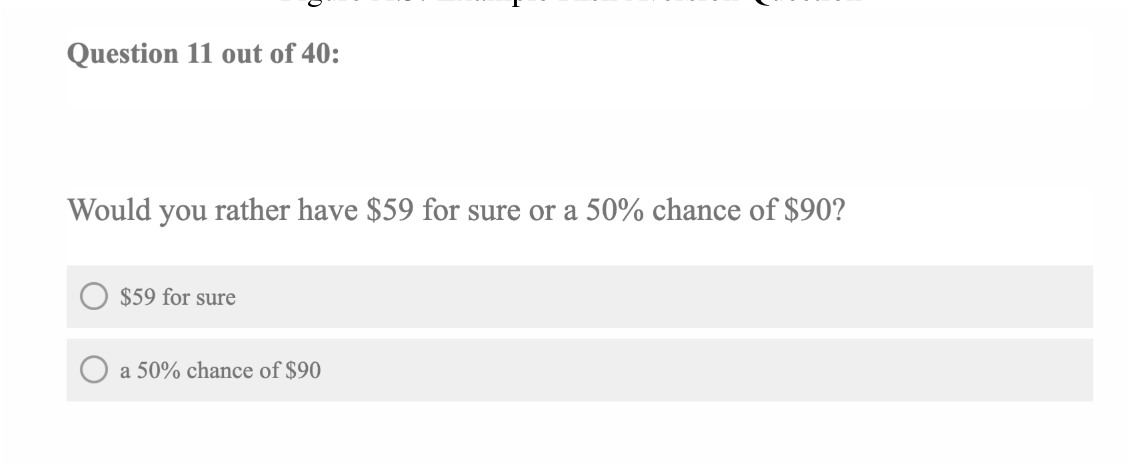
[1]The 95% confidence interval for the low incentive group for the effect of a $1 increase in incentives on WTP is $[0.008, 0.208]$ and the effect of reminders for this group is 0.29. For the high incentive group, the analogous calculations give $[-0.064, 0.464]$ and 0.16; thus, for this group, the confidence interval is too wide to permit a sharp test of the statement.

Table A.17: Effect of Risk Aversion (as Measured in Experiment 2 Part-2 Survey) on WTP

	WTP for Reminders ($)		
	(1)	(2)	(3)
Above Median Risky Choice Total	0.04		
	(0.085)		
Extra $1	0.08	0.07	0.05
	(0.025)	(0.030)	(0.030)
High Incentive	0.96	0.99	1.04
	(0.128)	(0.152)	(0.125)
Above Median Risky Choice Total × Extra $1	0.02		
	(0.042)		
Above Median Risky Choice Total × High Incentive	-0.15		
	(0.209)		
Extra $1 × High Incentive	-0.13	-0.20	0.04
	(0.077)	(0.093)	(0.083)
Above Median Risky Choice Total × Extra $1 × High Incentive	0.09		
	(0.117)		
Fraction of Risky Choices		0.09	
		(0.131)	
Fraction of Risky Choices × Extra $1		0.03	
		(0.070)	
Fraction of Risky Choices × High Incentive		-0.23	
		(0.310)	
Fraction of Risky Choices × Extra $1 × High Incentive		0.29	
		(0.189)	
Completed Part-2 Survey			-0.04
			(0.068)
Completed Part-2 Survey × Extra $1			0.04
			(0.037)
Completed Part-2 Survey × High Incentive			-0.14
			(0.165)
Completed Part-2 Survey × Extra $1 × High Incentive			-0.15
			(0.104)
Constant	0.49	0.48	0.53
	(0.054)	(0.064)	(0.052)
Observations	20,912	20,912	33,216
Number of Participants	1,307	1,307	2,076

This table estimates how willingness to pay for reminders varies with level of risk aversion and with completion of the survey. The risk aversion variables are derived from participant answers to 10 questions in the the part-2 survey of experiment 2, in which participants were asked to select between receiving a "for sure" amount of money and receiving a higher amount of money with 50% probability (see Figure A.5 for an example of how these questions were presented). The Above Median Risky Choice Total variable is an indicator for whether the number of times that a participant selected the uncertain choice is higher than the median. The Fraction of Risky Choices variable is the fraction of times a participant selected the uncertain choice. Columns (1) and (2) focus on the effect of risk-seeking behavior and are restricted to participants who were never top-coded at $3 or $11 and who were never top-coded at $4 or $12, as in all of the willingness-to-pay analysis. Since all participants included in Columns (1) and (2) completed part-2 of the survey, Column (3) focuses on the representativeness of willingness-to-pay responses for those who were randomized to receive or not receive reminders and who completed the part-2 survey. Standard errors, clustered at the participant level, are shown in parentheses.

Figure A.5: Example Risk Aversion Question



Question 11 out of 40:

Would you rather have $59 for sure or a 50% chance of $90?

- ◯ $59 for sure
- ◯ a 50% chance of $90

Participants answered 10 questions in this format, where the first value was randomly drawn from the integers between 35 and 65, and the second value was randomly drawn from the integers between 90 and 100.

References

CHETTY, RAJ (2006): "A New Method of Estimating Risk Aversion," *The American Economic Review*, 96, 1821–1834.

CHETTY, RAJ AND ADAM SZEIDL (2007): "Consumption commitments and risk preferences," *The Quarterly Journal of Economics*, 122, 831–877.

COHEN, ALMA AND LIRAN EINAV (2007): "Estimating Risk Preferences from Deductible Choice," *American Economic Review*, 97, 745–788.

HANDEL, BENJAMIN R (2013): "Adverse selection and inertia in health insurance markets: When nudging hurts," *The American Economic Review*, 103, 2643–2682.

HOLT, CHARLES A. AND SUSAN K. LAURY (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655.

MILGROM, PAUL AND ILYA SEGAL (2002): "Envelope Theorems for Arbitrary Choice sets," *Econometrica*, 70, 583–601.

SYDNOR, JUSTIN (2010): "(Over)insuring Modest Risks," *American Economic Journal: Applied Economics*, 2, 177–179.

VON GAUDECKER, HANS-MARTIN, ARTHUR VAN SOEST, AND ERIK WENGSTROM (2011): "Heterogeneity in Risky Choice Behavior in a Broad Population," *American Economic Review*, 101, 664–694.