Temptation and Self-Control (Gul and Pesendorfer)

Todd Sarver

Econ 885 – Duke University

Outline

Partial Commitments: A General Framework

- Motivation and Framework
- Strotz Representation
- Self-Control Representation
- From Self-Control to Strotz

Axiomatic Analysis

- Self-Control and Strotz Representation Results
- Dynamic Self-Control Preferences

Outline

Partial Commitments: A General Framework

- Motivation and Framework
- Strotz Representation
- Self-Control Representation
- From Self-Control to Strotz

Axiomatic Analysis

- Self-Control and Strotz Representation Results
- Dynamic Self-Control Preferences

The u-v Model

- Laibson (1997) showed that illiquid assets can be used to (partially) control future behavior when preferences exhibit quasi-hyperbolic discounting.

- Generalizing this idea, in this section we develop a simple, abstract approach to modeling partial commitments.

- Begin with what we'll call the "u-v model":
  - Consider a model with two periods: Periods 1 and 2.
  - Alternatives come from some set C.
  - Choice if could commit in period 1 maximizes $u(c)$.
  - Choice if made in period 2 instead maximizes $v(c)$.

- When full commitment is not possible, how do we model partial commitments made in period 1 to constrain behavior in period 2?

Example (Lunch)

- Imagine an individual will have the choice of the following options for lunch: salad (s), chicken sandwich (c), and hamburger (h).
- If she is able to commit to her lunch decision now, she ranks the alternatives according to their health attributes and hence would choose the salad.
- If she chooses at lunchtime, she is subject to a craving for richer foods and instead chooses the hamburger.
- These choices can be represented using the following utility functions (u, v):

$$u(s) = 5 > u(c) = 4 > u(h) = 1$$
$$v(s) = 3 < v(c) = 5 < v(h) = 6.$$

Example (Lunch, continued)

$$u(s) = 5 > u(c) = 4 > u(h) = 1$$
$$v(s) = 3 < v(c) = 5 < v(h) = 6.$$

- Suppose the individual cannot fully commit to her choice of lunchtime meal, but she can partially commit by choosing between restaurants with different menus.
- We can model different restaurants using different menus of options, for example,
$$\{c\} \quad \text{or} \quad \{s, c\} \quad \text{or} \quad \{s, c, h\}.$$
- Can we say something about the comparison of these (and other) option sets based on u and v?

Framework for Evaluating Option Sets (Partial Commitments)

- Assume for now that C is a finite set (we will work with more general sets of alternatives later).
- Let X be the set of all non-empty subsets of C.
  - We refer to $x \in X$ as an option set or menu.
- In period 1, the individual has a preference $\succsim_1$ on X.
  - For example, suppose the individual can choose between menus x and y in period 1. If $x \succ_1 y$, then the individual strictly prefers to choose x in period 1, and this becomes her choice set in period 2.
- In period 2, the individual has a preference $\succsim_2$ on C.
  - For example, if menu x was selected in period 1, then in period 2 chooses the alternative $c \in x$ that is most preferred according to $\succsim_2$.

From Commitment Rankings to Ranking Option Sets

- Suppose the individual ranks (full commitment to) alternatives in period 1 according to $u(c)$ and ranks alternatives in period 2 according to $v(c)$. What does this tell us about $\succsim_1$?
  - Clearly, u tells us how the individual would compare full-commitment menus:

$$\forall c, d \in C : \quad \{c\} \succsim_1 \{d\} \iff u(c) \geq u(d)$$

- But we would like to determine a value function $U(x)$ for evaluating all menus $x \in X$ in period 1, that is, a representation for $\succsim_1$.
  - For instance, how would the individual with u and v defined as in the last example rank the menus $\{s, c\}$ and $\{s, c, h\}$?
  - Or the menus $\{c\}$ and $\{s, c, h\}$?

- It turns out there are multiple possibilities, and we'll explore two: the Strotz Representation and the Self-Control Representation.

Outline

Strotz Representation

- The first approach we'll consider is based on the notion of consistent planning originally due to Strotz (1955): The individual chooses the best plan among those that she will actually follow, and evaluates menus accordingly.
- This is the approach/model that we have been implicitly using so far in the course, for example, when studying Laibson (1997) and Harris and Laibson (2001).

Strotz Representation

Definition (Strotz Representation)

A preference $\succsim_1$ has a Strotz representation if there is a pair of utility functions (u, v) defined on C such that $\succsim_1$ is represented by the value function $U : X \to \mathbb{R}$ defined by

$$U(x) = \max_{c \in x} u(c) \quad \text{subject to} \quad v(c) \geq v(d),\ \forall d \in x.$$

- Revisiting our example: $\{s, c\}$ versus $\{s, c, h\}$?
  - A sophisticated individual (we will discuss naivete later in the course) recognizes she will choose according to v in period 2, and therefore knows she will choose c from the first menu and h from the second.
  - Therefore, given her period 1 ranking of alternatives, she will prefer to choose the menu $\{s, c\}$ in period 1: $\{s, c\} \succ_1 \{s, c, h\}$.

Strotz Representation

Note that we can write the value function for the Strotz representation more succinctly as

$$U(x) = \max_{c \in B_v(x)} u(c)$$

where

$$B_v(x) = \operatorname*{argmax}_{c \in x} v(c).$$

If we are examining both stages of choice explicitly, then we can extend this definition as follows:

Definition (Two-Stage Strotz Representation)

The pair $(\succsim_1, \succsim_2)$ has a Strotz representation (u, v) if

1. $\succsim_1$ is represented by the function U defined above and
2. $\succsim_2$ is represented by lexicographic maximization of v then u. That is, $c \succsim_2 d$ if and only if

$$v(c) > v(d) \quad \text{or} \quad [v(c) = v(d) \text{ and } u(c) \geq u(d)].$$

Strotz Representation

$$U(x) = \max_{c \in B_v(x)} u(c)$$

The $B_v(x)$ constraint formalizes Strotz's "consistent planning":

- We can think of $B_v(x)$ as the result of imposing an additional constraint on the plans the period 1 self can implement, beyond the physical constraint set x.
- The period 1 self would like to choose the alternative in x that maximizes u, but she recognizes that she will not be able to follow through on a plan to choose some element $c \in x \setminus B_v(x)$.
- Instead, she must select an alternative from the set of options she will actually be willing to choose later, $B_v(x)$. Hence she engages in consistent planning.

Example

Example (Lunch, continued)

$$\begin{aligned} u(s) = 5 > u(c) = 4 > u(h) = 1 \\ v(s) = 3 < v(c) = 5 < v(h) = 6. \end{aligned}$$

Since $v(h) > v(c) > v(s)$ (i.e., $h \succ_2 c \succ_2 s$), the period 1 values for the possible subsets of $\{c, s, h\}$ are as follows:

$$\begin{aligned} U(\{s,c,h\}) = U(\{s,h\}) = U(\{c,h\}) = U(\{h\}) &= 1 \\ U(\{s,c\}) = U(\{c\}) &= 4 \\ U(\{s\}) &= 5. \end{aligned}$$

Thus

$$\begin{aligned} \{s\} &\succ_1 \{c\} \sim_1 \{s,c\} \\ &\succ_1 \{h\} \sim_1 \{s,h\} \sim_1 \{c,h\} \sim_1 \{s,c,h\}. \end{aligned}$$

The Laibson (1997) model is a recursive version of the Strotz representation. We can use a simple two-period version to illustrate.

Example (Laibson (1997))

- Suppose $C \subset \mathbb{R}^2$.
    - Denote the elements of this set by $c = (c_2, c_3) \in C$.
    - Thus the choices are over consumption in periods 2 and 3.
    - We suppress income and period 1 consumption for simplicity.
- Then, u and v are given by

$$
\begin{aligned}
u(c) &= \beta\delta u_2(c_2) + \beta\delta^2 u_3(c_3) \\
v(c) &= u_2(c_2) + \beta\delta u_3(c_3)
\end{aligned}
$$

  We could equivalently let $u(c) = u_2(c_2) + \delta u_3(c_3)$.
- Given period 1 wealth w_1, the amount $z_1 \geq 0$ allocated to the illiquid asset determines the period 2 constraint set:

$$
x^{z_1} = \left\{ (c_2, c_3) \in \mathbb{R}^2 : c_2 \leq R(w_1 - z_1),\ c_3 \leq R(Rw_1 - c_2) \right\}
$$

Outline

Partial Commitments: A General Framework

- Motivation and Framework
- Strotz Representation
- Self-Control Representation
- From Self-Control to Strotz

Axiomatic Analysis

- Self-Control and Strotz Representation Results
- Dynamic Self-Control Preferences

Temptation without Self-Control

- For the moment, we again restrict attention to a finite set of alternatives C for simplicity.
- Notice that in the Strotz representation, if $\{c\} \succ_1 \{d\}$ then there are only two possible period 1 rankings of $\{c, d\}$, depending on the choices in period 2:
  - $\{c\} \sim_1 \{c, d\} \succ_1 \{d\}$ if c will be chosen in period 2 ($c \succsim_2 d$).
  - $\{c\} \succ_1 \{c, d\} \sim_1 \{d\}$ if d will be chosen in period 2 ($d \succ_2 c$).

Temptation with Costly Self-Control

- Gul and Pesendorfer (2001) suggested that there are many instances in which individuals suffer from temptation yet are able to exert costly self-control to refrain from choosing the tempting alternative.
- In this case, if $\{c\} \succ_1 \{d\}$, the following rankings capture the different possibilities regarding temptation:
  - $\{c\} \succ_1 \{c, d\} \sim_1 \{d\}$: d is tempting, and the individual succumbs to temptation.
  - $\{c\} \succ_1 \{c, d\} \succ_1 \{d\}$: d is tempting, but the individual resists temptation at some psychological cost.
  - $\{c\} \sim_1 \{c, d\} \succ_1 \{d\}$: d is not tempting.

  The Strotz representation only permits the first and last ranking, but not the middle possibility.

Self-Control Representation

Definition (Self-Control Representation)

The preference $\succsim_1$ has a self-control representation if there is a pair of utility functions (u, v) defined on C such that $\succsim_1$ is represented by the value function $U : X \to \mathbb{R}$ defined by

$$U(x) = \max_{c \in x} \left[u(c) + v(c) \right] - \max_{d \in x} v(d).$$

- Writing in terms of self-control costs:

$$U(x) = \max_{c \in x} \left(u(c) - \underbrace{\left[\max_{d \in x} v(d) - v(c) \right]}_{\text{self-control cost } \mathbf{c}(c,x)} \right)$$

- Dual objectives: maximizing $u(c)$ and minimizing self-control costs. Optimal choice from menus compromises between these objectives by maximizing $u(c) + v(c)$:

$$\operatorname*{argmax}_{c \in x} \left(u(c) - \left[\max_{d \in x} v(d) - v(c) \right] \right) = \operatorname*{argmax}_{c \in x} \left[u(c) + v(c) \right]$$

Self-Control Representation

This interpretation suggests that period 2 choice should maximize $u + v$. We can define a two-period representation along these lines.

Definition (Two-Stage Self-Control Representation)

The pair $(\succsim_1, \succsim_2)$ has a self-control representation (u, v) if

1. $\succsim_1$ is represented by

$$U(x) = \max_{c \in x} \left[u(c) + v(c) \right] - \max_{d \in x} v(d)$$

2. $\succsim_2$ is represented by $u(c) + v(c)$. That is,

$$c \succsim_2 d \iff u(c) + v(c) \geq u(d) + v(d).$$

- In contrast to the Strotz representation, second period choice maximizes $u + v$ not v (but we'll see a connection momentarily).
- Note also that there is no need for lexicographic maximization of $u + v$ then u, since all maximizers give the same ex ante value.

Self-Control Representation

Summary

- Period 1 value for menus: $U(x) = \max_{c \in x} \left[u(c) + v(c)\right] - \max_{d \in x} v(d)$
- Period 1 value for full-commitment menus: $U(\{c\}) = u(c)$
- Period 2 ranking of alternatives: $u(c) + v(c)$

- Note the "non-consequentialist" evaluation of menus: unchosen options may adversely affect utility.
- GP draw a distinction between preference change and temptation:
    - Both the period 1 and period 2 self have the same objective u. They interpret this a stable preference, where it's simply that any choices made in the second period are subject to temptation and self-control costs. This perspective is useful for welfare analysis.
    - Since commitment choices in period 1 maximize $u(c)$ while choices in period 2 maximize $u(c) + v(c)$, the individual nonetheless exhibits dynamically inconsistent behavior.

Example (Lunch—now with self-control)

Recall previous numerical example, but now also need to calculate $u+v$:

	s	c	h
u	5	4	1
v	3	5	6
$u+v$	8	9	7

- $\{s\} \succ_1 \{s,h\} \succ_1 \{h\}$: exerts costly self-control and chooses s

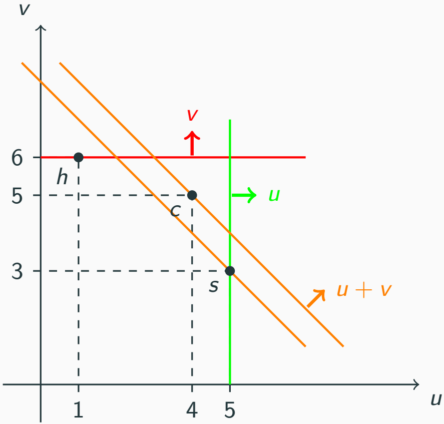
$$\begin{aligned} U(\{s\}) &= u(s) = 5 \\ U(\{h\}) &= u(h) = 1 \\ U(\{s,h\}) &= \max_{d \in \{s,h\}} \big[u(d)+v(d)\big] - \max_{d \in \{s,h\}} v(d) = 8-6 = 2 \end{aligned}$$

- $\{s\} \succ_1 \{s,c\} \sim_1 \{c\}$: succumbs to temptation to choose c

$$\begin{aligned} U(\{c\}) &= u(c) = 4 \\ U(\{s,c\}) &= \max_{d \in \{s,c\}} \big[u(d)+v(d)\big] - \max_{d \in \{s,c\}} v(d) = 9-5 = 4 \end{aligned}$$

Illustration in (u, v) space

	s	c	h
u	5	4	1
v	3	5	6
$u+v$	8	9	7



Outline

Magnitude of Self-Control Costs

- Note that in the Strotz representation, the scale of v doesn't matter, only the ranking it induces.
- In the GP representation, the scale of v matters: Replacing v with λv changes period 1 menu rankings and period 2 choices.
- Intuition: higher $\lambda \implies$ higher cost of self-control.

Example (Lunch—now with self-control, continued)

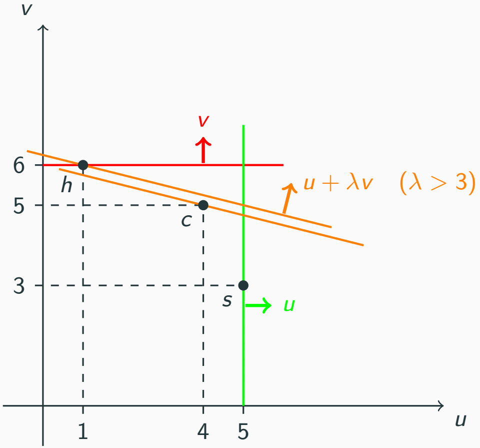
	s	c	h
u	5	4	1
v	3	5	6
$u + \lambda v$	$5 + \lambda 3$	$4 + \lambda 5$	$1 + \lambda 6$

For the self-control representation $(u, \lambda v)$:

- $h \succsim_2 s$ and $\{s, h\} \sim_1 \{h\}$ for $\lambda \geq 4/3$
- $h \succsim_2 c$ and $\{s, c, h\} \sim_1 \{c, h\} \sim_1 \{h\}$ for $\lambda \geq 3$.

Illustration in (u, v) space

	s	c	h
u	5	4	1
v	3	5	6



Strotz Representation as a Limiting Case

As the previous results example suggests, the Strotz representation is the limiting case of the self-control representation as the cost of exerting self-control (parameterized by λ) goes to ∞.

Lemma

Let U_λ denote the value function for the self-control representation $(u, \lambda v)$:

$$U_\lambda(x) = \max_{c \in x} \left[u(c) + \lambda v(c)\right] - \max_{d \in x} \lambda v(d).$$

Let U denoted the value function for the Strotz representation (u, v):

$$U(x) = \max_{c \in B_v(x)} u(c).$$

Then $U_\lambda(x) \to U(x)$ for all x as $\lambda \to \infty$.

Proof (for finite C).

Step 1 — show $U(x) \leq \lim U_\lambda(x)$: For any λ, the self-control representation always gives weakly higher utility than the Strotz representation:

$$U(x) = \max_{c \in B_v(x)} u(c) = \max_{c \in B_v(x)} \left[u(c) + \lambda v(c)\right] - \max_{d \in x} \lambda v(d) \leq U_\lambda(x),$$

Step 2 — show $U(x) \geq \lim U_\lambda(x)$:

- Since C is finite, there exists $\lambda' > 0$ such that if $c \notin B_v(x)$ then $c \notin B_{u+\lambda v}(x)$ for all $\lambda > \lambda'$. Thus, for $\lambda > \lambda'$ we have $B_{u+\lambda v}(x) \subset B_v(x)$. That is, any maximizer of $u + \lambda v$ is also a maximizer of v.
- Fix any $\lambda > \lambda'$ and any $c^* \in B_{u+\lambda v}(x)$. Since $c^* \in B_v(x)$, there are no self-control costs associated with choosing c^* and hence

$$\begin{aligned} U_\lambda(x) &= \max_{c \in x} \left[u(c) + \lambda v(c)\right] - \max_{d \in x} \lambda v(d) \\ &= \underbrace{\left[u(c^*) + \lambda v(c^*)\right]}_{c^* \in B_{u+\lambda v}(x)} - \underbrace{\lambda v(c^*)}_{c^* \in B_v(x)} = u(c^*) \leq \max_{c \in B_v(x)} u(c) = U(x). \end{aligned}$$

$\square$

Outline

Partial Commitments: A General Framework

- Motivation and Framework
- Strotz Representation
- Self-Control Representation
- From Self-Control to Strotz

Axiomatic Analysis

- Self-Control and Strotz Representation Results
- Dynamic Self-Control Preferences

Outline

Partial Commitments: A General Framework

- Motivation and Framework
- Strotz Representation
- Self-Control Representation
- From Self-Control to Strotz

Axiomatic Analysis

- Self-Control and Strotz Representation Results
- Dynamic Self-Control Preferences

Setting

Gul and Pesendorfer (2001) choice setting:

- C compact metric space.
- $p, q \in \triangle(C)$ — lotteries (Borel probability measures) over C.
- $x, y \in X \equiv \mathcal{K}(\triangle(C))$ — compact subsets of $\triangle(C)$.
- Primitive: Binary relation $\succsim$ on X.
    - No period 2 choice in our discussion of axioms.
    - See Lipman and Pesendorfer (2013) for a survey that includes axioms for the two-period setting discussed earlier.

Technical Details

- Mixture operation on X:

$$\alpha x + (1-\alpha)y \equiv \{\alpha p + (1-\alpha)q : p \in x, q \in y\}.$$

- Topological conditions:
  - C compact metric space.
  - $\triangle(C)$ is endowed with the topology of weak convergence. With this topology, it is compact and metrizable using the Prohorov metric (same as the Euclidean metric if C is finite).
  - $X \equiv \mathcal{K}(\triangle(C))$ is compact when endowed with the Hausdorff metric: If ρ denotes the (Prohorov) metric on $\triangle(C)$, the Hausdorff metric d_h on X is defined by

$$d_h(x,y) = \max\left\{\max_{p\in x}\min_{q\in y}\rho(p,q), \max_{q\in y}\min_{p\in x}\rho(p,q)\right\}.$$

Self-Control Representation

Definition

A relation $\succsim$ has a self-control representation if there exist continuous linear functions $u : \triangle(C) \to \mathbb{R}$ and $v : \triangle(C) \to \mathbb{R}$ such that the function $U : X \to \mathbb{R}$ defined by

$$U(x) = \max_{p \in x} \big[u(p) + v(p)\big] - \max_{q \in x} v(q)$$

for $x \in X$ represents $\succsim$.

- Note that continuous linear functions have expected-utility representations: $u(p) = \int \hat{u}(c)\, dp(c)$ and $v(p) = \int \hat{v}(c)\, dp(c)$.
- Can show that U is continuous by Berge's Maximum Theorem.

Axioms

- Axiom 1 (Weak Order): $\succsim$ is complete and transitive.
- Axiom 2 (Continuity): The sets $\{y \in X : y \succsim x\}$ and $\{y \in X : x \succsim y\}$ are closed for every $x \in X$.
- Axiom 3 (Independence): $x \succ y$ and $\alpha \in (0,1)$ implies $\alpha x + (1-\alpha)z \succ \alpha y + (1-\alpha)z$.
- Axiom 4 (Set Betweenness): $x \succsim y$ implies $x \succsim x \cup y \succsim y$.
    - This is the main substantive axiom for capturing temptation and self-control. Generalizes what we already observed for singleton menus.
    - The interpretation in terms of self-control is just as before. Recall:
        - $\{c\} \sim \{c, d\} \succ \{d\}$: d is not tempting.
        - $\{c\} \succ \{c, d\} \succ \{d\}$: exerts self-control to resist temptation.
        - $\{c\} \succ \{c, d\} \sim \{d\}$: succumbs to temptation.
    - More generally: $x \succ x \cup y \succ y \implies$ has self-control at $x \cup y$.

Self-Control Representation Theorem

The main result from Gul and Pesendorfer (2001) is the following.

Theorem

The relation $\succsim$ satisfies Axioms 1–4 if and only if it has a self-control representation (u, v).

- Exercise: Show necessity of axioms.

Strotz Representation

The definition in this domain is just as before:

Definition

A relation $\succsim$ has a Strotz representation if there exist continuous linear functions $u : \triangle(C) \to \mathbb{R}$ and $v : \triangle(C) \to \mathbb{R}$ such that the function $U : X \to \mathbb{R}$ defined by

$$U(x) = \max_{p \in B_v(x)} u(p)$$

for $x \in X$ represents $\succsim$, where $B_v(x) = \operatorname{argmax}_{q \in x} v(q)$.

- Notice (check!) that this representation satisfies Weak Order, Independence, and Set Betweenness.
- The only axiom it violates is Continuity.

Strotz Representation Theorem

- GP consider a set of continuity axioms (Axioms 2a–2c in the paper) that are jointly weaker than the Continuity Axiom above. Without getting into the technical details, let's call the combination of these weaker conditions Weak Continuity.

Theorem

The relation $\succsim$ satisfies Weak Order, Weak Continuity, Independence, and Set Betweenness if and only if it has either a Self-Control Representation or a Strotz Representation.

- Axiom 5 (No Self-Control): Either $x \sim x \cup y$ or $y \sim x \cup y$.
  - Note that No Self-Control $\implies$ Set Betweenness

Theorem

The relation $\succsim$ satisfies Weak Order, Weak Continuity, Independence, and No Self-Control if and only if it has a Strotz Representation.

Outline

Partial Commitments: A General Framework

- Motivation and Framework
- Strotz Representation
- Self-Control Representation
- From Self-Control to Strotz

Axiomatic Analysis

- Self-Control and Strotz Representation Results
- Dynamic Self-Control Preferences

Further Reading

- Recall that the self-control representation is continuous.
- Through a carefully crafted recursive application of this representation, one can therefore obtain a model that permits discontinuities in consumption but nonetheless has a continuous value function.
- I'll leave it to you to explore these models. A rough overview:
  - There is a brief introduction to these types of recursive models in the survey by Lipman and Pesendorfer (2013).
  - In terms of original research papers, a natural starting point is Gul and Pesendorfer (2004).
  - A useful generalization of their model was developed (non-axiomatically) by Krusell, Kuruşçu, and Smith (2010), and studied axiomatically by Noor (2011).
  - An extension of these models to permit naivete about future temptation was developed in Ahn, Iijima, and Sarver (2020).

Ahn, D. S., R. Iijima, and T. Sarver (2020): "Naivete About Temptation and Self-Control: Foundations for Recursive Naive Quasi-Hyperbolic Discounting", *Journal of Economic Theory*, 189, 105087 (page 39).

Gul, F. and W. Pesendorfer (2001): "Temptation and Self-Control", *Econometrica*, 69, 1403–1435 (pages 18, 31, 35).

Gul, F. and W. Pesendorfer (2004): "Self-Control and the Theory of Consumption", *Econometrica*, 72, 119–158 (page 39).

Harris, C. and D. Laibson (2001): "Dynamic Choices of Hyperbolic Consumers", *Econometrica*, 69, 935–957 (page 10).

Krusell, P., B. Kuruşçu, and A. A. Smith (2010): "Temptation and Taxation", *Econometrica*, 78, 2063–2084 (page 39).

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting", *Quarterly Journal of Economics*, 112, 443–478 (pages 4, 10, 15).

Lipman, B. L. and W. Pesendorfer (2013): "Temptation", in *Advances in Economics and Econometrics: Tenth World Congress*, ed. by D. Acemoglu, M. Arellano, and E. Dekel, vol. 1, Cambridge University Press (pages 31, 39).

Noor, J. (2011): "Temptation and Revealed Preference", *Econometrica*, 79, 601–644 (page 39).

Strotz, R. H. (1955): "Myopia and Inconsistency in Dynamic Utility Maximization", *Review of Economic Studies*, 23, 165–180 (page 10).