Revealing Naïveté and Sophistication from Procrastination and Preproperation

David J. Freeman*

June 9, 2020

Abstract

This paper proposes a novel way of distinguishing whether a person is naïve or sophisticated about her own dynamic inconsistency using only her task-completion behavior. It shows that adding an unused extra opportunity to complete a task can lead a naïve (but not a sophisticated) person to complete it later and can lead a sophisticated (but not a naïve) person to complete the task earlier. These results provide a framework for revealing preference and sophistication types from behavior in a general environment that includes that of O'Donoghue and Rabin (1999).

JEL Codes: D03, D84, D90.

Keywords: sophistication, naïveté, procrastination, preproperation, task completion, present bias, time inconsistency

*Address: Department of Economics, Simon Fraser University, 8888 University Dr, Burnaby, BC, Canada, V5A 1S6. E-mail: david_freeman@sfu.ca. I gratefully acknowledge funding from SFU VPR 4A Grant GRF-4A-2015-635001 from Simon Fraser University and SSHRC Insight Development Grant 430-2016-00193. I would like to thank David Ahn, Terri Kneeland, Kevin Laughren, Shih En Lu, Guy Mayraz, Luba Petersen, Philipp Sadowski, Todd Sarver, conference and seminar participants, and anonymous referees for numerous suggestions that improved this work. I also thank Harry David, Barbara Nordin, Matthew Pentecost, and Garrett Petersen for proofreading earlier versions of the manuscript.

Manuscript accepted at American Economic Journal: Microeconomics.

Behavioral models of intertemporal choice following Strotz (1955) incorporate two assumptions. First, a person may be dynamically inconsistent — that is, her current preferences over future actions may differ from the future preferences she will act on. Second, a person may be imperfectly self-aware of her own dynamic inconsistency when she forms expectations of her own future behavior. Strotz proposed two ways a dynamically inconsistent person might form expectations: she can be naïve and expect her future selves to behave according to her current preferences, or she can be sophisticated and hold correct expectations about her future behavior.

Yet neither a person's preferences nor her self-awareness is directly observed. Moreover, common domains of study, such as consumption-savings decisions, are insufficiently rich to jointly infer both intertemporal preferences and self-awareness (Blow, Browning and Crawford, 2017). This makes it difficult to understand which assumption (naïveté or sophistication) is more descriptively appropriate for any given application. In Strotz's model, however, a measure of self-awareness is necessary to accurately forecast the behavioral responses and welfare of an individual — for example, under different savings policies (Sprenger 2015, p. 283). This motivates the need to find ways to measure self-awareness from alternative domains of choice.[1]

This paper studies an individual's choice of when to complete a task that must be done exactly once, as in O'Donoghue and Rabin (1999). The paper's contribution is to show that sophistication and naïveté have sufficient force to yield distinct predictions in the domain of task completion. These predictions are directly testable and economically intuitive, and this is the case even in choice problems in which a person's higher-order beliefs about her future behavior are relevant and her preferences are not directly observed.

The main results of this paper establish that two separate classes of choice reversals are hallmarks of sophistication and naïveté. A person's behavior demonstrates

[1]Experiments that test for demand for commitment — a hallmark that indicates sophistication about time inconsistency — typically find that only a minority of participants do so (for example, Ashraf, Karlan and Yin (2006)). Section 7.4 reviews this literature. These findings further invite the question: if most people do not exhibit this hallmark of sophisticated time inconsistency, are most people naïve, or time consistent, or is something else going on?

a "reversal" if there is an instance in which adding an additional time period to a set of opportunities for completing a task changes the person's choice of when to complete the task, even though the additional opportunity will not be used. Reversals are distinguished between "doing-it-earlier" reversals, in which adding an opportunity leads the person to act earlier, and "doing-it-later" reversals, in which adding the opportunity leads them to act later in a sense that will be precisely defined. Theorems 2 and 3 establish that doing-it-later reversals are a hallmark of naïveté, while doing-it-earlier reversals are a hallmark of sophistication, in the sense that a time-inconsistent naïve person will exhibit a doing-it-later reversal while a sophisticate never will, and vice versa.

The following example shows how a doing-it-later reversal reveals naïveté. A student must do an assignment (act) exactly once in a given week, and we observe when she does it in each of two weeks. These weeks are directly comparable, in that the student's preferences over which day to do it are the same in both weeks. In the first week, she can only do it on either Tuesday or Thursday (because the lab is closed on Wednesday). In the second week, she also has the option of acting on Wednesday. Suppose we observe that she acts on Tuesday in the first week and Thursday in the second week. Her first week's behavior reveals that on Tuesday, she prefers to act now (that is, on Tuesday) rather than Thursday. With this observation in hand, her decision to wait on Tuesday in the second week reveals information about her preferences: on Tuesday, she must prefer to wait until Wednesday. It also provides information about her beliefs: she expects that she would act on Wednesday if she waits on Tuesday. Since she actually acts on Thursday in the second week, this expectation about her own behavior is incorrect, and she is revealed to be naïve about what she would do on Wednesday.

The next example shows how a doing-it-earlier reversal reveals sophistication. Suppose instead that the lab is open on Tuesday and Wednesday in the first week, that the lab is open those days plus Thursday in the second week, and that her preferences over which day to act are the same in each week. Suppose we observe that she does the assignment on Wednesday in the first week and Tuesday in the second week. Her first week's behavior reveals that on Tuesday she prefers to wait to act on Wednesday. With this observation in hand, her decision to act on Tuesday in the second week reveals information about her beliefs and preferences at that time: she expects that she would not act on Wednesday and she would prefer to act now (Tuesday) than on Thursday. She is thus not naïve about what she would do on Wednesday.

My main results show that the intuition from the above two examples can be extended more broadly. For a naïf who is not aware of her time inconsistency, adding an additional opportunity to act makes waiting appear weakly more attractive. Because of her time inconsistency, this can lead her to delay acting even if she does not use the added opportunity.

In contrast, a sophisticate correctly anticipates her future behavior; thus an additional opportunity will affect her earlier behavior only if she would act if she were to reach that opportunity. This limits the extent to which adding an additional opportunity can lead a sophisticate to delay. Moreover, if she knows she would act on this added opportunity against her earlier self's preferences, that earlier self may preemptively act. This can lead her to complete the task even earlier than without the added opportunity.

The results here concerning distinguishing naïfs from sophisticates relate to O'Donoghue and Rabin's (1999) findings that a naïve person will always act later than a sophisticated person with the same preferences and both will violate the Independence of Irrelevant Alternatives axiom. The results here also relate to O'Donoghue and Rabin's (2001) results that adding an additional completion opportunity can lead a partially naïve (but not a sophisticated) person to exhibit extreme procrastination. Unlike O'Donoghue and Rabin (1999; 2001), the formal choice environment here allows that the stream of payoffs associated with completing the task at a given point in time can be arbitrary and unobserved by the analyst. The treatment and main choice properties studied here apply regardless of whether the combination of the choice environment and the nature of dynamically inconsistent preferences induces future selves to want to delay (that is, procrastinate) or act early (that is, preproperate) against earlier selves' wishes or some combination of both — though stronger results are obtained in the first two cases. Furthermore, my results (unlike those of O'Donoghue and Rabin) do not make parametric assumptions about preferences.

The paper's results are as follows. The main results, Theorems 2 and 3, establish doing-it-later and doing-it-earlier reversals as hallmarks of naïveté and sophistication, respectively. Theorems 4 and 5 provide further restrictions on reversals for cases in which the form of time inconsistency tends to induce procrastination versus preproperation. Theorem 6 provides complete characterizations of naïveté and sophistication for the domain studied here, and Theorem 7 provides such a characterization for a model of partial naïveté. Corollary 1 and Proposition 2 provide conditions under which the relative sophistication of two partially naïve choice functions can be compared. Under the restrictions imposed by the task-completion domain here, Propositions 3-4 show that naïveté and sophistication each imply that behavior has an alternative representation in a well-studied two-stage model of boundedly rational choice, though each assumption implies a distinct model of the first stage (respectively, the models of Manzini and Mariotti 2007 and Masatlioglu, Nakajima and Ozbay 2012). Section 7 discusses possible extensions and reviews related literature.

1 Modeling naïveté and sophistication

1.1 Environment

Consider a person who faces a set of opportunities — periods in which a task can be done — and must act to complete a task in exactly one of the available periods. If the person has not already completed the task, she can either act (that is, complete the task) or wait. She cannot commit the behavior of her future selves except by acting.

Let $T \geq 3$ be a finite integer. Let $\bar{A} = \{1, \ldots, T\}$ denote the set of all possible periods; a period will typically be denoted by $t \in \bar{A}$. Let $\mathscr{A}$ denote the collection of all non-empty finite subsets of the set $\bar{A}$ of all possible opportunities. An opportunity set will typically be denoted by $A \in \mathscr{A}$.[2] Assume that a choice function $c : \mathscr{A} \to \bar{A}$ is observed, where $c(A)$ denotes the time a person acts when A is her

[2] Each $A \in \mathscr{A}$ is finite and thus induces a finite-horizon problem.

opportunity set.[3] While c is formally equivalent to a choice function on the domain $\mathscr{A}$ in the usual sense, its interpretation is different from that of a usual static choice function since it represents a dynamic choice problem and this is embedded in the temporal structure in $\bar{A}$. Interpret the behavior $t = c(A)$ as a result of choosing to wait at all periods in A prior to t and then choosing to act at time t rather than wait until later.

Given a set $A \in \mathscr{A}$ and $t \in \bar{A}$, let $A_{>t} = \{t' \in A : t' > t\}$ and define $A_{\geq t}$, $A_{<t}$, and $A_{\leq t}$ analogously.

Consider two examples that can fit into this framework.

Example 1. A statistics assignment must be done in a computer lab in a single sitting. The lab is only open on a subset of weekdays announced in advance. In this case, A could be given by any non-empty subset of $\{1,2,3,4,5\}$ and t denotes the opportunity to do the homework assignment on the t^{th} weekday.

Example 2. Let each $t \in \bar{A}$ be associated with a vector in $x^t \in \mathbb{R}^T$ that specifies a stream of costs/rewards associated with acting at time t. Setting $T = 2$ yields the setup of O'Donoghue and Rabin (1999), where $x^t = \left(x_1^t, x_2^t\right)$ means that acting at t yields an immediate utility benefit or cost of x_1^t at time t and a delayed cost or benefit of x_2^t realized at some time $\bar{T} > T$, where $\bar{T}$ does not depend on t.

Notice that the setup of this paper makes no distinction between choice environments with immediate costs and delayed rewards that are likely to induce procrastination versus choice environments with immediate benefits and delayed costs likely to invoke preproperation. Indeed, the setup here can include cases that do not fit either structure.

1.2 Preferences

Consider the following model of preferences that allows for changing tastes (following Strotz 1955). Each person has a set of time-dependent utility functions. For

[3] c is (technically) a choice function, rather than a correspondence, which rules out the possibility of genuine indifference. If indifference is always broken in the same deterministic way, this assumption is innocuous since T is finite.

each $t \in \bar{A}$, let $U_t : \bar{A}_{\geq t} \to \mathbb{R}$ denote her time-t utility function — that is, the utility function she uses at time t when she evaluates the desirability of each completion opportunity. Let $\mathscr{U} = \langle U_1, \ldots, U_T \rangle$ denote the ordered collection of utility functions for each t, and assume each is one-to-one.[4]

The structure of the completion-opportunity space and of people's preferences in this general model is minimally restricted. The magnitude and timing of flow utility the person expects to experience now or in the future (or both) if she acts in a particular period need not be observed by the analyst. This allows for arbitrary time-variant preferences.

1.3 Beliefs and behavior under time inconsistency

The behavior of a time-inconsistent person will depend both on her preferences (represented by $\mathscr{U}$) and on her expectations about her future behavior in each period.

Model a person's beliefs about her future behavior through a set of perceived future utility functions, where each such function captures an earlier period's beliefs about the utility function that will apply in a later period. For each $t_1, t_2 \in \bar{A}$ with $t_1 < t_2$, the function $\hat{U}_{t_2|t_1} : \bar{A}_{\geq t_2} \to \mathbb{R}$ denotes the utility function that time-t_1 self believes her time-t_2 self will apply, referred to as a perceived future utility function. Given any $t \in \bar{A}$, let $\hat{\mathscr{U}}_{\cdot|t} = \left\langle \hat{U}_{t+1|t}, \ldots, \hat{U}_{T|t} \right\rangle$ denote the ordered collection of the t-self's perceived future utility functions and let $\hat{\mathscr{U}} = \left\langle \hat{\mathscr{U}}_{\cdot|1}, \ldots, \hat{\mathscr{U}}_{\cdot|T-1} \right\rangle$ denote her ordered collection of all perceived future utility functions. This formulation rules out higher-order beliefs about future utility functions (perceived perceived future utility functions and so on): a person at time t forecasts that her future beliefs and preferences will both be determined by $\hat{\mathscr{U}}_{\cdot|t}$.

Given the preferences and beliefs captured by the pair $\left(\mathscr{U}, \hat{\mathscr{U}}\right)$, the perception-perfect equilibrium concept of O'Donoghue and Rabin specifies how a person will behave when facing any opportunity set. Let the function s denote a strategy, where $s(t, A, U_t, \hat{\mathscr{U}}_{\cdot|t}) =$ wait or $=$ act specifies a wait-or-act decision in period t in oppor-

[4] Since $\bar{A}$ is finite, this assumption rules out indifferences and is without loss of generality so long as we assume that all indifferences are broken in the same way.

tunity set A given a current utility function U_t and perceived future utility functions $\hat{\mathscr{U}}_{\cdot|t}$. A strategy is perception perfect if in each period t, the person best responds according to her utility function U_t to her beliefs about her future utility functions and about future beliefs determined by $\hat{\mathscr{U}}_{\cdot|t}$. Specifically, beliefs about behavior at future time t' are forecasted from predicted strategy $s(t', A, \hat{U}_{t'|t}, \hat{\mathscr{U}}_{\cdot|t})$, which is perception perfect for the utility functions she predicts she will apply in the future. Letting τ_t denote the period after t in which she predicts she first would act if she waits at t, she acts at t if $U_t(t) > U_t(\tau_t)$ and waits otherwise.

Definition. The <u>perception-perfect strategy</u> corresponding to opportunity set $A \in \mathscr{A}$, set of utility functions $\mathscr{U}$, and set of perceived future utility functions $\hat{\mathscr{U}}$ is a strategy s that, for each $t \in A$, satisfies:

$$s(t, A, U_t, \hat{\mathscr{U}}_{\cdot|t}) = \begin{cases} \text{act} & \text{if } U_t(t) > U_t(\hat{\tau}_t) \text{ or } A_{>t} = \emptyset \\ \text{wait} & \text{otherwise} \end{cases}$$

for $\hat{\tau}_t = \min\left\{\tau > t : s(\tau, A, \hat{U}_{\tau|t}, \hat{\mathscr{U}}_{\cdot|t}) = \text{act}\right\}$.

Next, introduce representations of behavior. A representation for a choice function is called a Strotzian representation if it models choice generated by a perception-perfect strategy for some sets of utility functions $\mathscr{U}$ and perceived future utility functions $\hat{\mathscr{U}}$. A Strotzian representation is (fully) naïve if every time-t self's perceived time-t' utility function is given by her current utility function U_t. In contrast, a Strotzian representation is sophisticated if every time-t self's perceived time-t' utility function is given by her actual time-t' utility function $U_{t'}$. These terms are defined formally below.

Definition. The choice function c has a <u>Strotzian</u> representation if there exist a $\mathscr{U}$ and a $\hat{\mathscr{U}}$ such that for each $A \in \mathscr{A}$, $c(A) = \min_t \{t : s(t, A, U_t, \hat{\mathscr{U}}_{\cdot|t}) = \text{act}\}$. A Strotzian representation is <u>naïve</u> if $\hat{U}_{t_2|t_1}(t_3) = U_{t_1}(t_3)$ for all $t_1, t_2, t_3 \in \bar{A}$ with $t_1 < t_2 \leq t_3$. A Strotzian representation is <u>sophisticated</u> if $\hat{U}_{t_2|t_1}(t_3) = U_{t_2}(t_3)$ for all $t_1, t_2, t_3 \in \bar{A}$ with $t_1 < t_2 \leq t_3$.

1.4 Time consistency and reversals

A person's preferences are time consistent if all of her utility functions are consistent with the same ranking over periods — that is, if no two periods' selves disagree on when to act. In the setting here, we only observe when the decision maker acts. Therefore, observationally, time consistency is equivalent to the requirement that the following two conditions hold for every opportunity set: (i) at all periods before she acts, she would rather wait until the time she actually acts and (ii) at the time she acts, she prefers acting to waiting until any available future period.

Definition. A choice function c is observationally time consistent if for every $A \in \mathscr{A}$, $t = c(A)$ implies that $t = c(\{t, t'\})$ for all $t' \in A$. Otherwise, c is observationally time inconsistent.

In the Strotzian model, completion times in two-opportunity choice sets reveal preferences at the earlier of the two periods. Thus, observational time consistency requires that each act-or-wait decision be made on the basis of a utility function that, as far as can be detected from observable choice, is consistent with the preferences of subsequent selves.

Example 3. Let $\bar{A} = \{1,2,3\}$, and suppose $3 = c(\{1,2,3\})$, $2 = c(\{1,2\})$, $3 = c(\{1,3\}) = c(\{2,3\})$. This c is observationally time consistent. Yet c has a Strotzian representation with $U_1(2) > U_1(3) > U_1(1)$, $U_2(3) > U_2(2)$, which would typically be viewed as a time-inconsistent representation. However, observational time inconsistency is a property of c, rather than $\mathscr{U}$. Notice that c also has a representation with $U_1(3) > U_1(2) > U_1(1)$, so choice alone cannot conclusively determine whether this c ought to be represented with a $\mathscr{U}$ that is time consistent in the usual sense; this motivates the notion of observational time consistency used in this paper.

The definition of observational time (in)consistency is based on comparing choice in arbitrary choice sets to choice in two-opportunity choice sets; the latter allow for clear inferences about preferences. However, the definition of observational time inconsistency does not indicate whether or how one can draw clear inferences about a person's beliefs about her own inconsistency. This motivates an alternative way to test for time inconsistency that will enable such inferences. To that end, introduce the notion of a reversal, based on a comparison of when a person does the task in two choice sets where one of them has an extra available opportunity.

Definition. c exhibits a reversal if there exists an $A \in \mathscr{A}$ and a t_2 such that $t_1 = c(A)$, $t_3 = c(A \cup \{t_2\})$, and $t_1, t_2 \neq t_3$. c exhibits no reversals if for all $A \in \mathscr{A}$ and $t_2 \in \bar{A}$, $c(A \cup \{t_2\}) \in \{c(A), t_2\}$.

The no-reversals property defined above is a variation of the Independence of Irrelevant Alternatives axiom and is equivalent to observational time consistency, as formalized below.[5]

Theorem 1. *Let c be a choice function. The following are equivalent: (i) c is observationally time consistent, (ii) c exhibits no reversals, and (iii) c has a Strotzian representation that is both sophisticated and naïve.*

Observational time inconsistency can thus be revealed from reversals (an insight that builds on O'Donoghue and Rabin 1999, pp. 114-115). The next section shows that specific classes of reversals allow an analyst to jointly infer time inconsistency and sophistication or naïveté about that inconsistency. However, when c is observationally time consistent, choices can be represented as maximizing a single preference relation or, equivalently, the same preferences in all periods. This holds if and only if sophisticated and naïve forecasts of future behavior coincide, as formalized in Theorem 1(iii).

2 Choice reversals under naïveté and sophistication

Consider two types of reversals a choice function might exhibit.

Definition. Consider a reversal with $t_1 = c(A)$, $t_3 = c(A \cup \{t_2\})$, and $t_1, t_2 \neq t_3$. The reversal is a doing-it-later reversal if $t_1 < t_3$ and either $t_2 < t_3$ or $t_1 = c(\{t_1, t_3\})$ (or both). The reversal is a doing-it-earlier reversal if $t_3 < t_1$.

[5] The no-reversals property is weaker than but equivalent to Sen's (1971, p. 313) Property α. Sen shows that his Property α is a necessary and sufficient condition for a (static) choice function on a finite domain to be rationalizable by an antisymmetric, complete, and transitive binary relation.

A doing-it-earlier reversal occurs when adding the period t_2 to A leads the person to act earlier than without the added opportunity. The definition of a doing-it-later reversal requires that a reversal satisfy both of two conditions. First, adding period t_2 to A results in the person's acting even later than without the added opportunity. Second,[6] either (i) her pairwise choice directly reveals that at $t_1 < t_3$, she prefers acting at t_1 over t_3 or (ii) she acts later in $A \cup \{t_2\}$ than the added period t_2 (or both i and ii). The results below show that naïve time inconsistency implies the existence of doing-it-later reversals and the absence of doing-it-earlier reversals. In contrast, sophistication allows doing-it-earlier but not doing-it-later reversals.

To see why a naïve person can exhibit doing-it-later reversals but not doing-it-earlier reversals, first notice that any person may find an added future opportunity attractive. However, a naïve person believes that if she currently most prefers that she act in that added future period, then she will continue to hold that preference in the future. However, this belief may be incorrect. If this belief fails to account for time inconsistency that would lead her future self to delay at her currently-most-preferred later period, or would lead a future self to act before that period, then she would exhibit a doing-it-later reversal. However, since a naïve person believes that her future behavior will be consistent with her current preferences, adding new opportunities to her opportunity set will make waiting appear weakly more attractive at earlier periods — thus a naïve person will never exhibit a doing-it-earlier reversal.

Example 4 gives an example of a doing-it-later reversal.

Example 4. Revisit Example 1. Suppose we observe that $1 = c(\{1,5\})$ and $5 = c(\{1,3,5\})$. These choices exhibit a doing-it-later reversal, since adding the unused option of acting on Wednesday (3, which is later than 1, Monday) leads the person to delay until Friday (5, which is later than Monday and Wednesday).

Example 5 gives an example of a doing-it-earlier reversal.

Example 5. Revisit Example 1. Suppose we observe that $3 = c(\{1,3\})$ and $1 = c(\{1,3,5\})$. Choices exhibit a doing-it-earlier reversal since adding the unused

[6]Example 6 illustrates why this second part of the definition of a doing-it-later reversal is needed for such reversals to reveal naïveté.

option to act on Friday ($t = 5$) leads the person to complete the assignment earlier than Wednesday.

Theorem 2 shows not only that it is possible that an added unused option will lead a naïf to delay but that any reversal exhibited by a naïf implies delay: she will never exhibit a doing-it-earlier reversal, and if she is time inconsistent she will exhibit at least one doing-it-later reversal.[7] An implication of this result is that the choices in Example 5 are inconsistent with naïve decision making, while those in Example 4 violate observational time consistency but are potentially consistent with naïve decision making.

Theorem 2. *If c has a naïve representation, then c does not exhibit any doing-it-earlier reversal; moreover, every reversal c exhibits is a doing-it-later reversal. If c is also observationally time inconsistent, then c exhibits a doing-it-later reversal.*

Proof. Suppose c has a naïve representation with set of utility functions $\mathscr{U}$ and set of perceived utility functions $\hat{\mathscr{U}}$. Further suppose c exhibits the reversal $t_1 = c(A), t_3 = c(A \cup \{t_2\}) \neq t_1, t_2$ and let $A' = A \cup \{t_2\}$. Let s denote a perception-perfect strategy.

Applying the restrictions that $\hat{U}_{t'|t} = U_t$ on $\bar{A}_{\geq t'}$ and that $A \subseteq A'$, it follows that for any $t < t_2$,

$$\begin{aligned} U_t(\tau_t) &= \max_{\tilde{t} \in A_{>t}} U_t(\tilde{t}) \\ &\leq \max_{\tilde{t} \in A'_{>t}} U_t(\tilde{t}) \\ &= U_t(\tau'_t) \end{aligned}$$

where $\tau_t = \min\{\tau > t : s(\tau, A, \hat{U}_{\tau|t}, \hat{\mathscr{U}}_{\cdot|t}) = \text{act}\}$ and $\tau'_t = \min\{\tau > t : s(\tau, A', \hat{U}_{\tau|t}, \hat{\mathscr{U}}_{\cdot|t}) = \text{act}\}$. Thus $s(t, A, U_t, \hat{\mathscr{U}}_{\cdot|t}) = \text{wait}$ implies $s(t, A', U_t, \hat{\mathscr{U}}_{\cdot|t}) = \text{wait}$ for such t. Thus $t_3 \geq \min\{t_1, t_2\}$.

First suppose $t_1 > t_2$. Then, $s(t, A', U_t, \hat{\mathscr{U}}_{\cdot|t}) = \text{wait}$ for all $t < t_2$. Since $c(A') \neq t_2$, we have $s(t_2, A', U_{t_2}, \hat{\mathscr{U}}_{\cdot|t_2}) = \text{wait}$ as well. Additionally, for all $t > t_2$ we have

[7] Theorem 2 here is related to Proposition 5 of O'Donoghue and Rabin (2001), which shows that for a person who faces the same set of options to complete a task each period, adding a new option to that set can lead to extreme procrastination if she is partially naïve, but not if she is sophisticated, under quasi-hyperbolic discounting. Neither result nests the other.

$A'_{>t} = A_{>t}$, from which it follows from the representation that $s(t, A', U_t, \hat{\mathscr{U}}_{\cdot|t}) = s(t, A, U_t, \hat{\mathscr{U}}_{\cdot|t})$. It thus follows that $t_3 = t_1$, which contradicts our initial assumption. Thus $t_1 > t_2$ cannot hold, and since $t_1 \neq t_2$, it follows that $t_2 > t_1$.

Since $t_3 \neq t_1$ and $t_3 \geq \min\{t_1, t_2\} = t_1$, we can conclude that $t_3 > t_1$. This establishes that a choice function with a naïve representation cannot exhibit a doing-it-earlier reversal.

Furthermore, since $t_1 = c(A)$ and $t_3 \in A_{>t_1}$, we must have $U_{t_1}(t_1) > U_{t_1}(\hat{\tau}_{t_1}) = \max_{t' \in A_{>t_1}} U_{t_1}(t') \geq U_{t_1}(t_3)$, which implies $t_1 = c(\{t_1, t_3\})$. Thus this reversal is a doing-it-later reversal. Since this reversal was selected arbitrarily, every reversal exhibited by c must be a doing-it-later reversal.

If c is observationally time inconsistent, then by Theorem 1 it exhibits at least one reversal and by the preceding argument this must be a doing-it-later reversal.

□

Turning to sophisticates, intuition suggests that sophisticates act earlier because they anticipate their future self-control problems. Thus they might act earlier to avoid the temptation to which they anticipate their future selves will succumb, thereby exercising the only type of commitment to which they have access in this choice environment. When adding a new opportunity leads a person to act at an earlier and previously available period, a person exhibits a doing-it-earlier reversal that cannot be accommodated by time-consistent preferences. Such behavior is, however, allowed for under time-inconsistent sophistication.

Theorem 3 states that a sophisticated choice function cannot exhibit a doing-it-later reversal but, if it is observationally time inconsistent, it exhibits at least one doing-it-earlier reversal.

Theorem 3. *If c has a sophisticated representation, then c does not exhibit any doing-it-later reversal. If c is also observationally time inconsistent, then c exhibits a doing-it-earlier reversal.*

Proof. Suppose c has a sophisticated representation with set of utility functions $\mathscr{U}$ and set of perceived utility functions $\hat{\mathscr{U}}$. Further suppose c exhibits the reversal $t_1 = c(A)$, $t_3 = c(A \cup \{t_2\}) \neq t_1, t_2$ and let $A' = A \cup \{t_2\}$. Let s denote the perception-perfect strategy.

Since $A_{>t_2} = A'_{>t_2}$, $s(t,A',U_t,\hat{\mathscr{U}}_{\cdot|t}) = s(t,A,U_t,\hat{\mathscr{U}}_{\cdot|t})$ for all $t > t_2$. Suppose $s(t_2,A',U_{t_2},\hat{\mathscr{U}}_{\cdot|t_2}) =$ wait. For $t_a = \max A_{<t_2}$, by sophistication, for each $t' > t_a$, $\hat{U}_{t'|t_2} = \hat{U}_{t'|t_a} = U_{t'}$ on $A_{\geq t'}$, so $s(t_2,A',\hat{U}_{t_2|t_a},\hat{\mathscr{U}}_{\cdot|t_a}) = s(t_2,A',U_{t_2},\hat{\mathscr{U}}_{\cdot|t_2}) =$ wait, and thus $\tau'_{t_a} = \min\{\tau > t_a : s(\tau,A',\hat{U}_{\tau|t_a},\hat{\mathscr{U}}_{\cdot|t_a}) = \text{act}\} = \min\{\tau > t_a : s(\tau,A,\hat{U}_{\tau|t_a},\hat{\mathscr{U}}_{\cdot|t_a}) = \text{act}\} = \tau_{t_a}$. By the same argument, if $t_b \in A_{<t_a}$ and $s(t,A',U_t,\hat{\mathscr{U}}_{\cdot|t}) = s(t,A,U_t,\hat{\mathscr{U}}_{\cdot|t})$ for all $t \in A$ with $t_b < t < t_a$, then $\tau'_{t_b} = \tau_{t_b}$ and thus $s(t_b,A',U_{t_b},\hat{\mathscr{U}}_{\cdot|t_b}) = s(t_b,A,U_{t_b},\hat{\mathscr{U}}_{\cdot|t_b})$. But then we have $s(t,A',U_t,\hat{\mathscr{U}}_{\cdot|t}) = s(t,A,U_t,\hat{\mathscr{U}}_{\cdot|t})$ for all $t \in A$. But this implies $t_3 = t_1$, a contradiction. Conclude that $s(t_2,A',U_{t_2},\hat{\mathscr{U}}_{\cdot|t_2}) = \text{act}$, and thus $t_3 < t_2$. Therefore c cannot exhibit a doing-it-later reversal with $t_3 > t_2$.

Now suppose that $t_1 < t_3$ and $t_1 = c(\{t_1,t_3\})$. By the representation, $t_1 = c(\{t_1,t_3\})$ implies $U_{t_1}(t_1) > U_{t_1}(t_3)$, and $t_3 = c(A \cup \{t_2\})$ implies that $t_3 = \min\left\{t : s(t,A',U_t,\hat{\mathscr{U}}_{\cdot|t}) = \text{act}\right\}$. But then by the definition of a perception-perfect strategy, $\text{act} = s(t_3,A',U_{t_3},\hat{\mathscr{U}}_{\cdot|t_3})$ and $\text{wait} = s(t,A',U_t,\hat{\mathscr{U}}_{\cdot|t})$ for all $t \in A_{<t_3} \cap A_{>t_1}$. Since $U_{t_1}(t_1) > U_{t_1}(t_3)$, we must also have $s(t_1,A',U_{t_1},\hat{\mathscr{U}}_{\cdot|t_1}) = \text{act}$, which would imply $c(A') \leq t_1$, a contradiction. Thus if c has a sophisticated representation, c cannot exhibit a doing-it-later reversal.

If c is observationally time inconsistent, then by Theorem 1 it exhibits at least one such reversal. If $t_1 > t_3$, then this is a doing-it-earlier reversal. Next, suppose $t_1 < t_3$. By the argument two paragraphs above, $s(t_2,A',U_{t_2},\hat{\mathscr{U}}_{\cdot|t_2}) = \text{act}$. Since $t_3 = c(A')$, it follows that $t_3 < t_2$, $s(t_3,A',U_{t_3},\hat{\mathscr{U}}_{\cdot|t_3}) = \text{act}$, and $s(t,A',U_t,\hat{\mathscr{U}}_{\cdot|t}) = \text{wait}$ for all $t \in A'_{<t_3}$. Working backward, if $s(t_3,A,U_{t_3},\hat{\mathscr{U}}_{\cdot|t_3}) = \text{act}$, then sophistication and $A_{<t_3} = A'_{<t_3}$ imply that $s(t,A,U_t,\hat{\mathscr{U}}_{\cdot|t}) = s(t,A',U_t,\hat{\mathscr{U}}_{\cdot|t}) = \text{wait}$ for all $t < t_3$ (including t_1), which contradicts that $t_1 < t_3$ and $t_1 = c(A)$; thus, $s(t_3,A,U_{t_3},\hat{\mathscr{U}}_{\cdot|t_3}) = \text{wait}$. By the sophisticated representation, it follows that $c(A_{\geq t_3}) > t_3$ and $c(A'_{\geq t_3}) = t_3$ is a doing-it-earlier reversal. □

Notice that a sophisticated choice function can still exhibit a reversal with $t_3 > t_1$ since not all such reversals are doing-it-later reversals; Example 6 illustrates such a case. However, Theorem 3 guarantees that if a sophisticated choice function c exhibits a reversal with $t_3 > t_1$, then both $t_3 = c(\{t_1,t_3\})$ and $t_3 < t_2$ must hold since c cannot exhibit a doing-it-later reversal.

Example 6. Revisit Example 1, and suppose that $1 = c(\{1,2,3\})$ but $2 = c(\{1,2,3,4\})$. If $2 = c(\{1,2\})$, this is not a doing-it-later reversal, since $2 < 4$. These choices would be generated by the sophisticated choice function with preferences $U_1(2) > U_1(1) > U_1(3) > U_1(4)$, $U_2(3) > U_2(2) > U_2(4)$, and $U_3(4) > U_3(3)$. Adding the opportunity to do the assignment on Thursday leads to delay compared to the case when the homework had to be completed by Wednesday. This occurs because the sophisticate waits on Monday because she recognizes that her Tuesday self will act then in order to avoid delaying until Thursday.

To summarize, Theorems 2 and 3 show that doing-it-later reversals are a hallmark of naïveté in the sense that a sophisticated person will not exhibit such reversals but a time-inconsistent naïve person will, whereas doing-it-earlier reversals are a hallmark of sophistication in the sense that a naïve person will never exhibit them but a time-inconsistent sophisticated person will.

3 Reversals for procrastination- vs. preproperation-inducing inconsistency

In many instances, it seems obvious a priori whether future selves will tend to wish to delay against the wishes of earlier selves or the reverse. In such cases, it is possible to more tightly characterize the form of reversals that naïve versus sophisticated people can exhibit.

To that end, introduce two classes of observationally time-inconsistent choice functions differentiated by the form of time inconsistency exhibited. A choice function suffers from "procrastination-inducing inconsistency" if each instance of observed time inconsistency involves a later self that would delay against the revealed preferences of an earlier self. Problems with immediate costs, delayed benefits, and present-biased preferences studied by O'Donoghue and Rabin (1999) exhibit procrastination-inducing inconsistency. Analogously, apply the term "preproperation-inducing inconsistency" to a choice function for which each instance of observed time inconsistency involves a later self that would complete the task earlier against the revealed preferences of an earlier self. Problems with immediate benefits, delayed costs, and present-biased preferences studied by O'Donoghue and Rabin (1999) exhibit preproperation-inducing inconsistency.[8] While these two types of problems are mutually exclusive, they are not exhaustive. In the definition of each, the restriction on observed time inconsistency only restricts cycles of choices from two-opportunity sets.[9] Thus they have implications for utility but not for beliefs in a Strotzian representation.

Definition. Periods t_1, t_2, t_3 form a cycle under c if $t_2 = c(\{t_1, t_2\})$, $t_1 = c(\{t_1, t_3\})$, and $t_3 = c(\{t_2, t_3\})$. A choice function exhibits only procrastination-inducing inconsistency if c is observationally time inconsistent and if for all t_1, t_2, t_3 with $t_1 < t_2, t_3$ that form a cycle, we have $t_2 < t_3$. A choice function exhibits only preproperation-inducing inconsistency if c is observationally time inconsistent and if for all t_1, t_2, t_3 with $t_1 < t_2, t_3$ that form a cycle, we have $t_2 > t_3$.

Notice that a naïve choice function that exhibits only procrastination-inducing inconsistency and exhibits cycle $t_2 = c(\{t_1, t_2\})$, $t_1 = c(\{t_1, t_3\})$, and $t_3 = c(\{t_2, t_3\})$ with $t_1 < t_2 < t_3$ will also have $t_3 = c(\{t_1, t_2, t_3\})$. This arises from the fact that, unanticipated by the t_1 self, the t_2-self waits, leading to procrastination. With preproperation-inducing inconsistency, a naïve choice function that exhibits this cycle will have $t_1 < t_3 < t_2$ and will also exhibit $t_3 = c(\{t_1, t_2, t_3\})$ — which is preproperation from the t_1-self's perspective since she wants her t_3-self to wait but the t_3-self instead acts. In this sense, procrastination- and preproperation-inducing inconsistency respectively induce procrastination and preproperation in simple three-opportunity choice sets for a naïve person.

To illustrate the definition, consider the following example of a choice function that exhibits procrastination-inducing inconsistency.

[8] A weaker sufficient condition for a choice function with a Strotzian representation to exhibit procrastination-inducing inconsistency is that $\mathscr{U}$ exhibits single-crossing differences in the sense that $t_1 < t_2 \leq t_3 < t_4$ and $U_{t_1}(t_4) > U_{t_1}(t_3)$ implies $U_{t_2}(t_4) > U_{t_2}(t_3)$. The analogous sufficient condition for preproperation-inducing inconsistency is $t_1 < t_2 \leq t_3 < t_4$ and $U_{t_1}(t_4) < U_{t_1}(t_3)$ implies $U_{t_2}(t_4) < U_{t_2}(t_3)$.

[9] Notice that any such cycle implies observational time inconsistency for any choice from the set containing all opportunities in the cycle. That is, if $t_2 = c(\{t_1, t_2\})$, $t_1 = c(\{t_1, t_3\})$, and $t_3 = c(\{t_2, t_3\})$, then any choice $c(\{t_1, t_2, t_3\})$ would violate observational time consistency.

Example 7. Consider a person who always wants to delay acting by one period, but never by more than that. That is, for each $t < T$, suppose $t+1 = c(\{t, t+1\})$ but $t = c(\{t, t+k\})$ for all $k > 1$. This choice function exhibits procrastination-inducing inconsistency. Notice that these binary choices imply that any Strotzian representation of c will have $U_t(t+1) > U_t(t) > U_t(t+k)$ for all $k \geq 2$.

Next consider the analogous example for preproperation-inducing inconsistency.

Example 8. Consider a person who always wants to act immediately rather than delay one period but would prefer waiting longer to acting immediately. That is, for each $t < T$, suppose $t = c(\{t, t+1\})$ but $t+k = c(\{t, t+k\})$ for all $k > 2$. This choice function exhibits preproperation-inducing inconsistency. Notice that these binary choices imply that any Strotzian representation of c will have $U_t(t+k) > U_t(t) > U_t(t+1)$ for all $k \geq 2$.

To provide a finer characterization of reversals demonstrated by a naïve person separately for choice functions exhibiting procrastination- and preproperation-inducing inconsistency, define the class of strong doing-it-later reversals as reversals in which the person acts later than the added but unused opportunity. That is, $t_1 = c(A)$ and $t_3 = c(A \cup \{t_2\}) > t_1, t_2$.

Definition. A reversal $t_1 = c(A)$, $t_3 = c(A \cup \{t_2\}) \neq t_1, t_2$ is a strong doing-it-later reversal if $t_3 > t_2, t_1$.

Where the definition of a doing-it-later reversal requires either $t_2 < t_3$ or $t_1 = c(\{t_1, t_3\})$, a strong doing-it-later requires the former. However, every strong doing-it-later reversal exhibited by a naïve choice function also satisfies the latter property.

Proposition 1. *If c is naïve and exhibits the strong doing-it-later reversal $t_1 = c(A)$, $t_3 = c(A \cup \{t_2\}) > t_1, t_2$, then $t_1 = c(\{t_1, t_3\})$ as well.*

It is now possible to state tighter restrictions on the type of reversal that a naïve person can exhibit separately for the special cases of choice functions that exhibit procrastination- and preproperation-inducing inconsistency. Under procrastination-inducing inconsistency, all reversals exhibited by a naïve person are strong doing-it-later reversals. Under preproperation-inducing inconsistency, a naïve person never exhibits strong doing-it-later reversals; in this case, all reversals are doing-it-later reversals that satisfy both $t_1 < t_3 < t_2$ and $t_1 = c(\{t_1, t_3\})$.

Theorem 4. *(i) If c is naïve and exhibits only procrastination-inducing inconsistency, then every reversal c exhibits is a strong doing-it-later reversal.*

(ii) If c is naïve and exhibits only preproperation-inducing inconsistency, then c does not exhibit any strong doing-it-later reversal.

It is also possible to obtain stronger restrictions on reversals that can be exhibited by sophisticated choice functions, particularly in the cases of procrastination- and preproperation-inducing inconsistency. For every reversal exhibited by a sophisticated choice function, adding an unused opportunity always leads to acting before that unused opportunity. A sophisticated choice function that exhibits only procrastination-inducing inconsistency can only exhibit a reversal in which the added unused opportunity is later than the period in which the person would have acted without it. Under preproperation-inducing inconsistency, the only restriction is that both the unused added opportunity and the completion time when it is added are either both earlier or both later than the original completion time.

Theorem 5. *If c is sophisticated and exhibits reversal $t_1 = c(A)$, $t_3 = c(A \cup \{t_2\}) \neq t_1, t_2$, then*

(i) $t_3 < t_2$,

(ii) if c exhibits only procrastination-inducing inconsistency, then $t_2 > t_1$,

(iii) if c exhibits only preproperation-inducing inconsistency, then either $t_3 < t_2 < t_1$ or $t_1 < t_3 < t_2$.

4 Characterizations of naïve and sophisticated choice

This section provides testable conditions that completely characterize the behavioral content of naïve and sophisticated choice in this domain.

Example 9 shows that without assuming additional structure, the absence of doing-it-later reversals does not guarantee that choice has a sophisticated representation.

Example 9. Consider a choice function c with a Strotzian representation with $U_1(2) > U_1(1) > U_1(3) > U_1(4), U_2(3) > U_2(2) > U_2(4), U_3(4) > U_3(3), \hat{U}_{3|1}(3) > \hat{U}_{3|1}(4)$, and $\hat{U}_{2|1} = U_2$, $\hat{U}_{3|2} = U_3$. In this representation, the time-1 self is naïve about her time 3 behavior, but all selves are otherwise sophisticated. We can see that the c corresponding to this $\left(\mathscr{U}, \hat{\mathscr{U}}\right)$ pair exhibits no doing-it-later reversals and does exhibit doing-it-earlier reversals since, applying the perception-perfect strategy, $2 = c(\{1,2,4\})$ but $1 = c(\{1,2,3,4\})$, and also $3 = c(\{2,3\})$ but $2 = c(\{2,3,4\})$. However, a fully sophisticated choice function $\tilde{c}$ with the same $\mathscr{U}$ would have $2 = \tilde{c}(\{1,2,3,4\})$. Since U_2 and U_3 are pinned down by choices from two-option sets, it follows that c exhibits no doing-it-later reversals yet does not have a sophisticated representation.

Next, consider three conditions that will be used to characterize naïve and sophisticated choice.

First, consider the Irrelevant Alternatives Delay condition, which rules out doing-it-earlier reversals while also requiring that the added option (t_2) in any reversal be at a later time than that of the initial choice (t_1).

Irrelevant Alternatives Delay. If $t_1 = c(A)$, and $t_1, t_2 \neq t_3 = c(A \cup \{t_2\})$, then $t_2, t_3 > t_1$.

The following property, Exclusion Consistency (Rubinstein and Salant, 2008),[10] specifies the restriction that if adding a period (t_2) generates a reversal, then the previously chosen option (t_1) is not chosen in any set in which the added period (t_2) is available.

Exclusion Consistency. If $t_1 = c(A)$, and $t_1, t_2 \neq c(A \cup \{t_2\})$, and $t_2 \in A'$, then $t_1 \neq c(A')$.

The Recursivity condition strengthens the central postulate of sophisticated consistent planning (Strotz, 1955; Pollak, 1968) as a choice-based condition. It requires

[10] Rubinstein and Salant (2008) use the Exclusion Consistency property to characterize a model of a person who chooses by sequentially applying two different binary relations. Their characterization uses an abstract choice space that lacks the temporal structure used here, and the relationship between models is discussed in Section 6.

that the added earlier option is taken if and only if it would be taken in the opportunity set that contains only two periods: the added earlier opportunity and the opportunity that would have been taken were the added option not available.[11]

Recursivity. For any A and t, $c(A_{>t} \cup \{t\}) = c(\{t, c(A_{>t})\})$.

Recursivity characterizes sophistication in this domain, while Irrelevant Alternatives Delay and Exclusion Consistency jointly characterize naïveté.

Theorem 6. *(i) c has a naïve representation if and only if c satisfies Irrelevant Alternatives Delay and Exclusion Consistency.*

(ii) c has a sophisticated representation if and only if c satisfies Recursivity.

These characterization results provide a complete set of testable implications of naïve and sophisticated representations. A full proof is available in the Appendix. Notice that beliefs about future behavior are trivial in opportunity sets with only two periods, and thus behavior in such sets cleanly reveals preferences that applied at the earlier period. A key step of the proof uses binary choices to characterize preferences: $t = c(\{t, t'\})$ implies $U_{\min\{t,t'\}}(t) > U_{\min\{t,t'\}}(t')$. Then, the proof shows that the conditions in either (i) or (ii) are sufficient to guarantee that $\mathscr{U}$ can be so constructed as to be consistent with naïveté and sophistication, respectively.

The following example shows that Irrelevant Alternatives Delay is not sufficient to guarantee that a choice function has a naïve representation.

Example 10. Let $T = 3$, $1 = c(\{1,2\}) = c(\{1,3\})$ and $2 = c(\{2,3\}) = c(\{1,2,3\})$. By construction, c satisfies Irrelevant Alternatives Delay. But in a naïve representation $1 = c(\{1,2\}) = c(\{1,3\})$ can occur only if $U_1(1) > U_1(2), U_1(3)$, which would imply $1 = c(\{1,2,3\})$ regardless of expectations about $t = 2$ behavior.

5 Partial naïveté

Some people may be neither fully naïve nor fully sophisticated. This section studies the behavior of such people. It will first, characterize the predictive content of the

[11] This property plays a similar role to Gul and Pesendorfer's (2005) "No Compromise" axiom and Noor's (2011) "Sophistication" axiom.

class of Strotzian and partially naïve representations in terms of two behavioral conditions, as explained below. Then, it will discuss how to compare the relative sophistication of two choice functions.

Define the term "partially naïve" to denote Strotzian representations in which all incorrect forecasts about future preferences align with current preferences. That is, a Strotzian representation is termed partially naïve if every time-t self's perceived time-t' utility function's ranking between t' and a later period is consistent either with her actual time-t' preferences or with her time-t preferences between these two periods. Sophisticated and naïve representations are thus each special cases of a partially naïve representation.

Definition. A Strotzian representation $\left(\mathscr{U}, \hat{\mathscr{U}}\right)$ is partially naïve if, for all $t_1, t_2, t_3 \in \bar{A}$ with $t_1 < t_2, t_3$, $U_{t_1}(t_2) > U_{t_1}(t_3)$ and $U_{\min\{t_2,t_3\}}(t_2) > U_{\min\{t_2,t_3\}}(t_3)$ implies $\hat{U}_{\min\{t_2,t_3\}|t_1}(t_2) > \hat{U}_{\min\{t_2,t_3\}|t_1}(t_3)$.

Not all Strotzian representations are partially naïve. For example, the Strotzian representation with utility functions $U_1(2) > U_1(1) > U_1(3)$ and $U_2(2) > U_2(3)$ and perceived future utility function $\hat{U}_{2|1}(3) > \hat{U}_{2|1}(2)$ is not partially naïve.

It may not be obvious what restrictions the classes of Strotzian and partially naïve representations place on behavior. To clarify, consider the following approach to revealing beliefs, which motivates characterizations of each representation. Suppose t_{n-1} and t_n are the last two available opportunities to do the task in choice set A and t_1 is an earlier opportunity in A. Then, a decision maker's time-t_1 decision should be invariant to replacing the last two periods, t_{n-1} and t_n, with the period in which she expects (at t_1) that she would first act if she were to reach the penultimate opportunity t_{n-1}. This reasoning applies regardless of the decision maker's sophistication. By this logic, if the time-t_1 choice depends on which period is removed, then for every other choice set A' where t_{n-1} and t_n are the last two available opportunities and t_1 is in A', the decision maker would hold the same t_1 beliefs about when she would act were she to reach t_{n-1}. The Penultimate Replaceability condition imposes this as a restriction on behavior. For each $t < T$, define the function $c_t : \{A \in \mathscr{A} : c(A) \leq t\} \to \{\text{wait}, \text{act}\}$ by $c_t(A) = \text{wait}$ if $c(A) > t$ and $c_t(A) = \text{act}$ if $c(A) = t$.

Penultimate Replaceability. Let $t_1 < t_2 < t_3$. Either $c_{t_1}(A \cup \{t_2, t_3\}) = c_{t_1}(A \cup \{t_2\})$ for all $A \subseteq \{t_1, t_1+1, \ldots, t_2-1\}$, or $c_{t_1}(A \cup \{t_2, t_3\}) = c_{t_1}(A \cup \{t_3\})$ for all $A \subseteq \{t_1, t_1+1, \ldots, t_2-1\}$.

Notice that by working backward, Penultimate Replaceability implies the following, stronger condition, End Replaceability (stated and proven as Proposition 5 in the Appendix).

End Replaceability. Let $t_1 < t_2 < \cdots < t_n$ and $A' = \{t_2, \ldots, t_n\}$. There exists a $t_k \in A'$ such that $c_{t_1}(A \cup A') = c_{t_1}(A \cup \{t_k\})$ for all $A \subseteq \{t_1, t_1+1, \ldots, t_2-1\}$.

When Penultimate Replaceability holds, the option that can be replaced for t_{n-1} and t_n will, in the representation, be the time at which the person believes (at time t_1) she would end up acting if she reached the penultimate opportunity, t_{n-1}. The condition places no restriction on when beliefs can be incorrect. But in a partially naïve representation, a person at time t_1 would only make an incorrect prediction about her t_{n-1} behavior by incorrectly applying t_1 preferences that differ from t_{n-1} preferences. When such behavior identifies incorrect t_1 expectations about what she would do if she reached t_{n-1}, if c has a partially naïve representation this also identifies the earlier self's preferences. The Wishfulness condition requires that for each earlier period t, time-t preferences can be constructed to align with any revealed mispredictions about behavior in the last two periods without generating any preference cycles. For each $t \in \bar{A}$, define R_t on $\bar{A}_{\geq t}$ as follows: if $t = c(\{t, t_1\})$, then $t R_t t_1$, if $t_1 = c(\{t, t_1\})$, then $t_1 R_t t$, and if there exists an $A \subseteq \{t, \ldots, \min\{t_1, t_2\} - 1\}$ such that $c_t(A \cup \{t_1, t_2\}) = c_t(A \cup \{t_1\}) \neq c_t(A \cup \{t_2\})$ but $t_2 = c(\{t_1, t_2\})$, then $t_1 R_t t_2$.

Wishfulness. For each $t \in \bar{A}$, R_t is acyclic.

Theorem 7 shows that choice has a Strotzian representation if and only if c satisfies Penultimate Replaceability, and this is a partially naïve representation if and only if c also satisfies Wishfulness.

Theorem 7. *(i) c has a Strotzian representation if and only if c satisfies Penultimate Replaceability.*

(ii) c has a partially naïve representation if and only if c satisfies Penultimate Replaceability and Wishfulness.

Different partially naïve people may fall at different places on the continuum between "naïve" and "sophisticated." This raises the question of how to compare the degree of sophistication of different Strotzian representations using behavior.

The intuition from the Penultimate Replaceability condition suggests a way of inferring earlier beliefs from behavior in a Strotzian representation. Beliefs at t_1 about time-t_2 preferences between t_2 and $t_3 > t_2$ can be revealed by looking at cases in which t_2 and t_3 are the last two opportunities. In such cases (by Penultimate Replaceability), behavior at t_1 is invariant to eliminating either t_2 or t_3. However, depending on t_1's preferences and beliefs about subsequent preferences, it is possible that for every such choice set, t_1 behavior is the same regardless of whether t_2 or t_3 is eliminated. For example, consider someone with the preferences $U_1(2), U_1(3) > U_1(1)$. Her time-1 beliefs about time-2 preferences between acting at time 2 vs. time 3 do not affect her behavior in opportunity set $\{1,2,3\}$. Thus, these beliefs cannot be revealed from choice. A comparison of the degree of sophistication between two people is only meaningful when comparing beliefs that are revealed from behavior.

To enable such a behavioral comparison, we can restrict comparisons to representations that are minimally naïve in the sense that when beliefs cannot be determined from behavior using the approach outlined above, they are modeled as sophisticated.

Definition. A partially naïve representation $\left(\mathscr{U}, \hat{\mathscr{U}}\right)$ for choice function c is <u>minimally naïve</u> if $c_{t_1}(A \cup \{t_2, t_3\}) = c_{t_1}(A \cup \{t_2\}) = c_{t_1}(A \cup \{t_3\})$ for all $A \subseteq \{t_1, \ldots, \min\{t_2, t_3\} - 1\}$ and $t_2 = c(\{t_2, t_3\})$ implies $\hat{U}_{\min\{t_2,t_3\}|t_1}(t_2) > \hat{U}_{\min\{t_2,t_3\}|t_1}(t_3)$.

Next, consider a working definition of what it means for one partially naïve choice function to be more sophisticated than another. Consider two partially naïve representations $\left(\mathscr{U}, \hat{\mathscr{U}}\right)$ and $\left(\mathscr{U}, \hat{\mathscr{U}}'\right)$, each with the same collection of utility functions $\mathscr{U}$ but with different forecasts of future utility functions at each period, captured by $\hat{\mathscr{U}}$ and $\hat{\mathscr{U}}'$. Intuitively, $\hat{\mathscr{U}}$ is more sophisticated than $\hat{\mathscr{U}}'$ if, whenever the perceived utility function in $\hat{\mathscr{U}}'$ makes a correct forecast about future behavior, so does the corresponding perceived utility function in $\hat{\mathscr{U}}$.[12]

Definition. Given two choice functions, c and c', with partially naïve representations $\left(\mathscr{U}, \hat{\mathscr{U}}\right)$ and $\left(\mathscr{U}, \hat{\mathscr{U}}'\right)$, respectively, c is more sophisticated than c' if for all $t_1, t_2, t_3 \in \bar{A}$ with $t_1 < t_2, t_3$, $U_{\min\{t_2,t_3\}}(t_2) > U_{\min\{t_2,t_3\}}(t_3)$ and $\hat{U}'_{\min\{t_2,t_3\}|t_1}(t_2) > \hat{U}'_{\min\{t_2,t_3\}|t_1}(t_3)$ implies $\hat{U}_{\min\{t_2,t_3\}|t_1}(t_2) > \hat{U}_{\min\{t_2,t_3\}|t_1}(t_3)$.

The "more sophisticated than" definition applies to two representations rather than their choices directly. The following comparison leverages intuition from the Penultimate Replaceability condition to compare two choice functions based on penultimate accuracy (defined below). When an earlier self's expectations about when she would act in the last two periods can be revealed from her behavior, these expectations may, but need not, match her actual action if the last two periods were reached. This gives a notion of penultimate accuracy that can be used to compare the sophistication of two choice functions.

Definition. c is more penultimately accurate than c' if for all $t_1, t_2, t_3 \in \bar{A}$ with $t_1 < t_2 < t_3$ and $A \in \mathscr{A}$ such that $A \subseteq \{t_1, \ldots, t_2 - 1\}$, $t_j = c'(\{t_2, t_3\})$ and $c'_{t_1}(A \cup \{t_2, t_3\}) = c'_{t_1}(A \cup \{t_j\}) \neq c'_t(A \cup \{t_{-j}\})$ implies $c_{t_1}(A \cup \{t_2, t_3\}) = c_{t_1}(A \cup \{c(\{t_2, t_3\})\})$.

The following corollary to Theorem 7 clarifies the tight link between the "more sophisticated than" relationship between representations and the "more penultimately accurate than" relationship between choice functions when restricted to minimally naïve representations.

Corollary 1. *Let c and c' have minimally naïve representations $\left(\mathscr{U}, \hat{\mathscr{U}}\right)$ and $\left(\mathscr{U}, \hat{\mathscr{U}}'\right)$. c is more sophisticated than c' if and only if c is more penultimately accurate than c'.*

The penultimately accurate relation is not obviously related to the reversals studied earlier. Example 11 shows why a simple comparison of reversals does not necessarily rank the relative sophistication of two choice functions. It shows that a

[12] The comparative-sophistication notion below is a restricted ordinal version of Ahn et al.'s (2019) "more u-aligned" notion for comparing perceived future utility functions, restricted to comparing individuals with identical preferences (that is, same $\mathscr{U}$).

partially naïve person can exhibit a doing-it-earlier reversal not exhibited by a fully sophisticated person with the same preferences. This is because believing that behavior will better align with current preferences can work in both directions. In the partially naïve model here, if a decision maker is currently more optimistic about her behavior in the more distant future, then she also expects her less-distant-future selves to share that optimism. This could lead her to expect future selves to delay against her current wishes, which could lead her to act earlier.

Example 11. O'Donoghue and Rabin (1999) show that a naïf will always act weakly later than a sophisticate with the same preferences facing the same choice set. This result does not transfer to a comparison between sophistication and partial naïveté. Consider $\bar{A} = \{1,2,3,4\}$ and $\mathscr{U}$ such that $U_1(2) > U_1(1) > U_1(3) > U_1(4)$, $U_2(3) > U_2(2) > U_2(4)$, and $U_3(4) > U_3(3)$. A sophisticate with these preferences has $2 = c(\{1,2,4\}) = c(\{1,2,3,4\})$. However, a partially naïve person who, at $t = 1$, incorrectly predicts that $\hat{U}_{3|1}(3) > \hat{U}_{3|1}(4)$ but correctly predicts that $\hat{U}_{2|1}(4) < \hat{U}_{2|1}(2) < \hat{U}_{2|1}(3)$ will expect herself to delay at $t = 2$ and act at $t = 3$ and would thus act at $t = 1$. This leads to a doing-it-earlier reversal since it implies $2 = c(\{1,2,4\})$ but $1 = c(\{1,2,3,4\})$. Thus for this $\mathscr{U}$, adding $t = 3$ to the opportunity set $\{1,2,4\}$ generates a doing-it-earlier reversal for this partially naïve person but not for a fully sophisticated one.[13] This demonstrates that the intuition that a more sophisticated person is more prone to doing-it-earlier reversals must be qualified.

It can also be shown that this partially naïve person exhibits no doing-it-later reversals in this domain. But the lack of a doing-it-later reversal that reveals her naïveté can be viewed as an artifact of the restrictive domain of this example: if acting at $t = 1$ could be made less attractive so that $U_1(3) > U_1(1) > U_1(4)$, then a doing-it-later reversal would be observed.

As shown in Example 11, the interaction between elements of sophistication and of naïveté can confound a clear comparison of degrees of sophistication by a simple comparison of their propensity for reversals when a person needs to form

[13]This example also demonstrates that a more sophisticated person need not complete the task earlier than a less sophisticated person with the same preferences.

beliefs about future beliefs. When a person faces at three completion opportunities, she must forecast only one future decision. In this case, such interactions do not arise, which allows for clear inferences about beliefs from behavior. This motivates a definition of "three-opportunity revealable" which delineates the class of beliefs that can be directly revealed by only looking at choices involving three or fewer opportunities.

Definition. A belief $\hat{U}_{\min\{t_2,t_3\}|t_1}(t_2) \gtrless \hat{U}_{\min\{t_2,t_3\}|t_1}(t_3)$ is three-opportunity revealable if $t_1 < \min\{t_2,t_3\}$ and $t_1 = c(\{t_1,t_2\}) \neq c(\{t_1,t_3\})$.

Next, compare the reversals involving only three options.

Definition. Let c and c' have partially naïve representations. Say that c exhibits more three-opportunity doing-it-later reversals than c' if for every triple $t_1,t_2,t_3 \in \bar{A}$ with $t_1 < t_2,t_3$, $t_1 = c'(\{t_1,t_3\}) = c(\{t_1,t_3\})$ and $t_3 = c'(\{t_1,t_2,t_3\})$ implies $t_3 = c(\{t_1,t_2,t_3\})$. Say that c exhibits more three-opportunity doing-it-earlier reversals than c' if for every triple $t_1,t_2,t_3 \in \bar{A}$ with $t_1 < t_2,t_3$, $t_3 = c'(\{t_1,t_3\}) = c(\{t_1,t_3\})$ and $t_1 = c'(\{t_1,t_2,t_3\})$ implies $t_1 = c(\{t_1,t_2,t_3\})$.[14]

Proposition 2 clarifies the relationship between these notions of exhibiting more three-opportunity doing-it-later/earlier reversals and the relative sophistication of two representations compared only for revealable beliefs.

Proposition 2. *Let c and c' have partially naïve representations $\left(\mathscr{U},\hat{\mathscr{U}}\right)$ and $\left(\mathscr{U},\hat{\mathscr{U}}'\right)$ respectively. The following are equivalent:*

(i) For any t_1,t_2,t_3 such that $\hat{U}_{\min\{t_2,t_3\}|t_1}(t_2) \gtrless \hat{U}_{\min\{t_2,t_3\}|t_1}(t_3)$ is three-opportunity revealable and $U_{\min\{t_2,t_3\}}(t_2) > U_{\min\{t_2,t_3\}}(t_3)$, then $\hat{U}_{\min\{t_2,t_3\}|t_1}(t_2) > \hat{U}_{\min\{t_2,t_3\}|t_1}(t_3)$ implies $\hat{U}'_{\min\{t_2,t_3\}|t_1}(t_2) > \hat{U}'_{\min\{t_2,t_3\}|t_1}(t_3)$;

(ii) c exhibits more three-opportunity doing-it-later reversals than c';

(iii) c' exhibits more three-opportunity doing-it-earlier reversals than c.

[14] Since picking an earlier action commits one to a decision, exhibiting more three-opportunity doing-it-earlier reversals is a special case of Gul and Pesendorfer's (2001) "greater preference for commitment" restricted to sets with up to three options. In a similar vein, exhibiting more three-opportunity doing-it-later reversals is a special case of Ahn et al.'s (2016) notion of being "more naïve" in the domain here.

An implication of Proposition 2 is that if all beliefs were three-opportunity revealable, then comparing only three-opportunity reversals would provide a complete basis for comparing the relative sophistication of two choice functions. Such an assumption would be reasonable in a sufficiently richer domain.[15]

6 Relationship to models of boundedly rational choice

In the setting considered in this paper, naïve and sophisticated representations have alternative representations in terms of boundedly rational choice procedures that have been previously studied. This section derives these connections. As a preliminary to the analysis below, given any binary relation R define $m(\cdot,R):\mathscr{A}\to\mathscr{A}$ by $m(A,R)=\{t\in A:\nexists t'\in A \text{ for which } t'Rt\}$.

First, introduce the rational shortlist method (RSM) of Manzini and Mariotti (2007).

Definition. A choice function c has an <u>RSM representation</u> if there exist asymmetric binary relations P_1 and P_2 such that $c(A)=m(m(A,P_1),P_2)$ for all $A\in\mathscr{A}$.[16]

Naïve choice behavior is equivalent to that of an RSM representation in which P_2 is the "less than" order $<$ on $\mathbb{N}$ restricted to $\bar{A}$ and P_1 is a subrelation of $>$.

Proposition 3. *c has a naïve representation if and only if it also has an RSM representation with $P_1\subseteq>$ and $P_2=<$.*

Rubinstein and Salant (2008) prove that a choice function c has an RSM representation if and only if it satisfies the Exclusion Consistency condition. Thus Proposition 3 can be viewed as a corollary of the conjunction of Rubinstein and Salant (2008) and Theorem 6. The proof here explicitly constructs P_1 and P_2 from $\mathscr{U}$ by providing an alternative characterization of a naïve representation.

[15] A working paper version of this paper (Freeman, 2016) proved such a result using a richer domain.

[16] Excuse the notational sloppiness in this definition that conflates each singleton set $m(m(A,P_1),P_2)$ with its element.

Proof. Suppose c has a naïve representation $\left(\mathscr{U},\hat{\mathscr{U}}\right)$. Then naïveté requires $\hat{U}_{t_2|t_1}(t_3) = U_{t_1}(t_3)$ for all t_1,t_2,t_3 with $t_1 < t_2 \leq t_3$. Plugging this into the definition of a predicted strategy and working backward from the last period with available opportunities in A, we obtain the result that at each t_1,t_2 with $t_1 < t_2$, $\text{act} = s(t_2,A,\hat{U}_{t_2|t_1},\hat{\mathscr{U}}_{\cdot|t_1})$ if and only if $t_2 = \underset{t\in A_{\geq t_2}}{\arg\max}\, U_{t_1}(t)$. Similarly, plugging the above formula for forecasts of behavior into the definition of s, we obtain an expression for $s(t,A,U_t,\hat{\mathscr{U}}_{\cdot|t})$:

$$s(t,A,U_t,\hat{\mathscr{U}}_{\cdot|t}) = \begin{cases} \text{act} & \text{if } U_t(t) > \max\limits_{t'\in A_{>t}} U_t(t') \\ \text{wait} & \text{otherwise} \end{cases}$$

This yields the representation that $c(A)$ is the earliest time at which that period's self prefers doing the task immediately to each available future opportunity. Let P_1 be a binary relation that ranks an opportunity above another if it prevents this doability condition from being satisfied. That is, define P_1 by $t'P_1t$ if and only if $t' > t$ and $U_t(t') > U_t(t)$; by construction, $P_1 \subseteq >$. Thus the set $m(A,P_1) = \{t \in A : s(t,A,U_t,\hat{\mathscr{U}}_{\cdot|t}) = \text{act}\}$ captures periods in which, if reached, the decision maker would act. Then, by construction, $m(m(A,P_1),<)$ returns the earliest time in A at which $s(t,A,U_t,\hat{\mathscr{U}}\cdot|t) = \text{act}$.

Conversely, suppose that c has an RSM representation P_1,P_2 where $P_1 \subseteq >$ and $P_2 = <$. Then define $\mathscr{U}$ so that $U_{\min\{t,t'\}}(t) > U_{\min\{t,t'\}}(t')$ if not $t'P_1t$. For any t and $t' > t$, define $\hat{U}_{t'|t} = U_t$ on $A_{\geq t'}$ and let s denote a perception-perfect strategy. By construction of $\mathscr{U}$ and $\hat{\mathscr{U}}$, given any set A, $m(A,P_1) = \{t \in A : U_t(t) > U_t(t')\ \forall t' \in A_{>t}\} = \{t \in A : s(t,A,U_t,\hat{\mathscr{U}}_{\cdot|t}) = \text{act}\}$, thus $m(m(A,P_1),P_2) = \min\{t \in A : s(t,A,U_t,\hat{\mathscr{U}}_{\cdot|t}) = \text{act}\}$. This establishes that this RSM representation has an alternative naïve representation. □

Just as any naïve choice function has an RSM representation, so too can any sophisticated representation be represented in a model of the boundedly-rational choice: the model of choice with limited attention (CLA) of Masatlioglu, Nakajima and Ozbay (2012).

Definition. A function $\Gamma : \mathscr{A} \to \mathscr{A}$ is an attention filter if, for each $A \in \mathscr{A}$, $\emptyset \neq$

$\Gamma(A) \subseteq A$ and for all $t \in A \backslash \Gamma(A)$ we have $\Gamma(A \backslash \{t\}) = \Gamma(A)$. A choice function c has a CLA representation if there exists an asymmetric, complete, and transitive binary relation P and an attention filter Γ such that $c(A) = m(\Gamma(A), P)$ for all $A \in \mathscr{A}$.

Proposition 4. *If c has a sophisticated representation, then it has a CLA representation with $P =<$.*

The proof explicitly constructs Γ and P from $\mathscr{U}$.

Proof. Let c have a sophisticated representation corresponding to $\mathscr{U}$. Define Γ by $\Gamma(A) = \{t \in A : s(t, A_{\geq t}, U_t, \hat{\mathscr{U}}_{\cdot|t}) = \text{act}\}$ for each A. That is, $\Gamma(A)$ gives the set of all periods in A in which, if reached, the person would act. Let $P =<$. Then $m(\Gamma(A), P)$ picks out the earliest period in A at which the strategy acts.

It remains to verify that Γ is an attention filter. Suppose $t \in A \backslash \Gamma(A)$. Then by our choice of Γ, $s(t, A_{\geq t}, U_t, \hat{\mathscr{U}}_{\cdot|t}) = \text{wait}$. But then $s(t', A_{\geq t'} \backslash \{t\}, U_{t'}, \hat{\mathscr{U}}_{\cdot|t'}) = s(t', A_{\geq t'}, U_{t'}, \hat{\mathscr{U}}_{\cdot|t'})$ for all $t' \neq t$: the $t' < t$ case follows by the definition of a sophisticated representation because $s(t, A_{\geq t}, U_t, \hat{\mathscr{U}}_{\cdot|t}) = \text{wait}$, and the $t' > t$ case follows trivially since $A_{\geq t'} = (A \backslash \{t\})_{\geq t'}$. □

The converse of Proposition 4 does not hold: the restrictions on Γ imposed by a CLA representation need not allow one to construct a $\mathscr{U}$ that generates the choices in a sophisticated representation. The following example shows that even a choice function with a CLA representation that exhibits no doing-it-later reversals can fail to have a sophisticated representation.

Example 12. Consider $\bar{A} = \{1, 2, 3, 4\}$. Let $\Gamma(\bar{A}) = \{1, 2, 3, 4\}$, $\Gamma(A) = A \backslash \{1\}$ for all $A \subsetneq \bar{A}$ with $A \neq \{1\}$, and let $P =<$. The choice function c with this CLA representation has $1 = c(\bar{A})$ and also has $2 = c(\{2, t\})$ for all t. By the latter, any sophisticated representation for c must have $U_1(2) > U_1(1)$ and $U_2(2) > U_2(3), U_2(4)$. These restrictions on U_1 and U_2 are sufficient to infer that 2 must be chosen from $\bar{A}$ in any sophisticated representation for c, which would contradict $1 = c(\bar{A})$. Thus c has no sophisticated representation.

7 Extensions and relation to existing work

7.1 Structured domains

The theoretical literature on time-inconsistent preferences studies individual preferences over dated rewards (Ok and Masatlioglu, 2007; Dziewulski, 2018; Ericson and Noor, 2015; Chakraborty, 2016), over consumption streams (Montiel Olea and Strzalecki, 2014; Noor and Takeoka, 2017; Galperti and Strulovici, 2017; Echenique, Imai and Saito, Forthcoming), and over lotteries over consumption streams (Hayashi, 2003).[17] These rich and structured domains allow these authors to study specific models of weighing earlier-versus-later rewards or consumption. However, these papers do not address how a decision maker resolves conflicting preferences when she cannot commit her future behavior. My setting is exactly suited to this task.

Also, unlike the aforementioned papers on time preferences, my setting does not assume that the analyst observes consumption or rewards realized in each period following an action. My model also does not impose structural assumptions on preferences. In contrast, most papers on time-inconsistent choice assume that preferences are time invariant (see Halevy 2015) and study particular functional forms for discounting current versus future payoffs (for example, quasi-hyperbolic discounting). That said, if the stream of payoffs associated with acting in each period were observable, variations on doing-it-later/earlier reversals could be generated to create new testable implications of naïveté and sophistication. This is illustrated with two short examples.

Suppose that a person makes choices from sets of dated rewards in $\mathbb{R} \times T$, where (x,t) indicates an opportunity at time t that yields immediate reward x. Suppose that the person always prefers larger rewards to smaller ones; that is, suppose each $U_t((x,t'))$ is increasing in x and $U_t((0,t')) = 0$. Let $0 < x \leq y < z' < y' < z$, and assume $(z,3) = c(\{(x,1),(y,2),(z,3)\})$ but $(x,1) = c(\{(x,1),(y',2),(z,3)\})$.

[17] Another example of a structured domain is that of optimal stopping with deterministic or stochastic payoff flows, as studied by Quah and Strulovici (2013). In Section V of their paper, they show that under partially naïve quasi-hyperbolic discounting, a more sophisticated agent will stop weakly earlier in their environment.

Such behavior is not a doing-it-earlier reversal, but it has the same intuition: in $\{(x,1),(y,2),(z,3)\}$, y is sufficiently low that the person would wait at $t=2$ and anticipate this at $t=1$, but y' is sufficiently high that when facing $\{(x,1),(y',2),(z,3)\}$ the person anticipates at $t=1$ that she would act at $t=2$ and avoids that by acting immediately. As in a doing-it-earlier reversal, a naïve person would never exhibit such a reversal, since she would only expect at $t=1$ that she would wait at $t=2$ if her $t=1$ self preferred to wait until $t=3$.

Now instead suppose that $(x,1) = \tilde{c}(\{(x,1),(y,2),(z',3)\})$ but $(y,2) = \tilde{c}(\{(x,1),(y,2),(z,3)\})$. The latter choice reveals that at $t=2$ she would act with $(y,2)$ over $(z,3)$ when facing $\{(y,2),(z,3)\}$. She would thus also act with $(y,2)$ over $(z',3)$ when facing $\{(y,2),(z',3)\}$. Thus if she were sophisticated, her waiting at $t=1$ when facing $\{(x,1),(y,2),(z,3)\}$ reveals that at $t=1$ she prefers $(y,2)$ over $(x,1)$. Such a preference, combined with sophistication, would be inconsistent with her acting at $t=1$ by doing $(x,1)$ when facing $\{(x,1),(y,2),(z',3)\}$. However, this pattern of behavior is consistent with a naïve specification of preferences and beliefs: if the $t=1$ self prefers $(z,3)$ over $(x,1)$ and prefers $(x,1)$ over $(y,2)$ and $(z',3)$, and naïvely expecting to wait at $t=2$ when facing $\{(x,1),(y,2),(z,3)\}$.

Many prior experiments study pairwise choices between dated rewards to estimate intertemporal utility functions. The examples above show that observing behavior in a modified domain can be similarly used to estimate a person's degree of sophistication.

7.2 Limited datasets

The doing-it-later and doing-it-earlier reversals each only apply in a restricted set of comparisons, and the characterizations in Theorem 6 assume that the analyst observes how a person would behave in all possible opportunity sets $A \in \mathscr{A}$. Yet the analysis here broadly points to testable content of both naïveté and sophistication that could be harnessed even when only a small number of choices are observed that may not allow direct tests for reversals. When only a subset of choice problems is observed, the no-reversals property (which is equivalent to the Weak Axiom of Revealed Preference in this domain) is insufficient to guarantee that choices will maximize a complete and transitive binary relation, motivating the Strong Axiom of Revealed Preference (Houthakker, 1950). The Online Appendix provides analogous solutions for testing naïveté and sophistication. It proposes tests that check whether it is possible to construct a set $\{R_t\}_{t=1}^T$ of complete, transitive, and asymmetric binary relations corresponding to $\mathscr{U}$ that generate a naïve/sophisticated representation consistent with observed choices. These tests are in the spirit of limited-dataset tests of models of boundedly rational choice outlined by De Clippel and Rozen (2018).

7.3 Relation to work in decision theory on choice over menus

Work on decision theory has used a person's explicit preferences over choice sets to characterize the implications of sophistication for people with time-inconsistent preferences (Gul and Pesendorfer, 2005) or intrinsic preferences for smaller choice sets due to self-control costs (Noor, 2011). These papers have not sought analogous characterizations of naïve choice. The comparison can be made more explicit by nesting the choice environment of this paper within that of Gul and Pesendorfer (2005); such an extension also allows Theorem 3 to be extended to sophisticated models with intrinsic self-control costs (like the model of Gul and Pesendorfer (2001; 2004)). In recent work in a two-period choice setting, Ahn et al. (2019) provide a way of comparing the relative naïveté of decision makers using their preferences over choice sets of lotteries and their subsequent choice. Their main result relates their comparative measure to properties of the $t = 1$ and $t = 2$ utility functions of expected utility maximizers, and they use the rich domain of lotteries to ensure that the extent of an individual's sophistication can be elicited. Since a person's higher-order beliefs about behavior are potentially relevant whenever a person faces a decision in more than two periods, none of their results directly apply to my setting when $T > 3$.

7.4 Relation with empirical work on measuring sophistication

Many existing experiments look for whether subjects demand commitment — a hallmark that indicates sophistication about a self-control problem. Experiments that are often cited as finding evidence of demand for commitment typically find that only 25–40 percent of participants demand commitment (Ashraf, Karlan and Yin, 2006; Kaur, Kremer and Mullainathan, 2010; Duflo, Kremer and Robinson, 2011; John, 2019). Augenblick, Niederle and Sprenger (2015) is the notable exception that further demonstrates the robustness of this finding: while 59 percent of their subjects demand commitment at a price of $0, only 9 percent of subjects are willing to pay a positive price for it.

A related line of work takes as a hallmark of (partial) naïveté a person's willingness to commit to a contract that leads to an ex post outcome that is strictly suboptimal according to ex ante preferences. Indeed, a doing-it-later reversal can be viewed as an example of such behavior. In another example, Giné, Karlan and Zinman (2010) find that of the (small minority of) smokers who took up a commitment savings account tied to passing a subsequent nicotine test, only 19 percent of those tested passed the nicotine test and the remainder lost all of their savings in that account. Heidhues and Kőszegi (2010) show that (partial or full) naïveté could lead a borrower to choose a contract that offers a cheap front-loaded repayment schedule with a high late-payment penalty and then end up paying the penalty, thereby achieving a worse outcome for the contract-signing self than she would have achieved from initially choosing an alternative back-loaded repayment. DellaVigna and Malmendier (2004; 2006) find that many people purchase gym memberships at a price that is cost-inefficient compared to a pay-per-use price given ex post use, and they similarly interpret their findings as evidence of partial naïveté.

Two other prior attempts to estimate a degree of sophistication estimate that people are, on average, completely naïve. But these papers' estimates require exceedingly strong assumptions. Fang and Wang (2015) assume that demographic heterogeneity related to one's expectation of contracting cancer is unrelated to degree of sophistication. Augenblick and Rabin (2019) assume that subjects' experimentally elicited beliefs about their future behavior correspond to the beliefs subjects act on when making real-world decisions. In my view, this growing evidence that suggests that people are truly naïve motivates the need for choice-based hallmarks of naïveté (such as doing-it-later reversals) and measures based on them.

Other previous experiments study the domain of task completion (Ariely and

Wertenbroch, 2002; Burger, Charness and Lynham, 2011; Bisin and Hyndman, 2014) but do not attempt to use their data to distinguish naïveté from sophistication.

8 Discussion

This paper has established separate hallmarks of naïveté and sophistication in the domain of task completion. The examples and results here suggest how empirical work on task completion can be used to measure naïveté versus sophistication based on behavior and how related results can be obtained for models of partial naïveté. The results of such tests would provide a much-needed means of evaluating the appropriateness of alternative assumptions about sophistication and naïveté in applications of models of time-inconsistent preferences.

One implication of the analysis here is that firms that have data on task completion can learn the degree of sophistication or naïveté of their clients or employees, especially if they can experiment. For example, a financial institution that observes when a client pays her bills can use this information to target her with financial products that exploit her degree of naïveté without having to offer her a menu of contracts that screen for this — targeting that Heidhues and Kőszegi (2017) show can lower welfare. Similarly, a manager who observes when an employee completes work assignments can use this information to infer the degree of sophistication of the employee and can use this to better tailor her work responsibilities and deadlines.

References

Ahn, David S, Ryota Iijima, Yves Le Yaouanq, and Todd Sarver. 2019. “Behavioural Characterizations of Naivete for Time-Inconsistent Preferences.” *Review of Economic Studies*, 86(6): 2319–2355.

Ariely, Dan, and Klaus Wertenbroch. 2002. “Procrastination, deadlines, and performance: Self-control by precommitment.” *Psychological Science*, 13(3): 219–224.

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. “Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines.” *Quarterly Journal of Economics*, 121(2): 635–672.

Augenblick, Ned, and Matthew Rabin. 2019. “An Experiment on Time Preference and Misprediction in Unpleasant Tasks.” *Review of Economic Studies*, 86(3): 941–975.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger. 2015. “Working Over Time: Dynamic Inconsistency in Real Effort Tasks.” *Quarterly Journal of Economics*, 130(3): 1067–1115.

Bisin, Alberto, and Kyle Hyndman. 2014. “Procrastination, Self-Imposed Deadlines and Other Commitment Devices.”

Blow, Laura, Martin Browning, and Ian Crawford. 2017. “Never mind the hyperbolics: nonparametric analysis of time-inconsistent preferences.” Working Paper.

Burger, Nicholas, Gary Charness, and John Lynham. 2011. “Field and online experiments on self-control.” *Journal of Economic Behavior & Organization*, 77(3): 393–404.

Chakraborty, A. 2016. “Present Bias.” Working Paper.

De Clippel, Geoffroy, and Kareen Rozen. 2018. “Bounded rationality and limited datasets.” Working Paper.

DellaVigna, Stefano, and Ulrike Malmendier. 2004. “Contract design and self-control: Theory and evidence.” *Quarterly Journal of Economics*, 119(2): 353–402.

DellaVigna, Stefano, and Ulrike Malmendier. 2006. "Paying not to go to the gym." *American Economic Review*, 96(3): 694–719.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2011. "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya." *American Economic Review*, 101(6): 2350–2390.

Dziewulski, Paweł. 2018. "Revealed time-preference." *Games and Economic Behavior*, 112: 67–77.

Echenique, Federico, Taisuke Imai, and Kota Saito. Forthcoming. "Testable Implications of Models of Intertemporal Choice: Exponential Discounting and Its Generalizations." AEJ: Microeconomics.

Ericson, Keith Marzilli, and Jawwad Noor. 2015. "Delay Functions as the Foundation of Time Preference: Testing for Separable Discounted Utility." Working Paper.

Fang, Hanming, and Yang Wang. 2015. "Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions." *International Economic Review*, 56(2): 565–596.

Freeman, David. 2016. "Revealing Naïveté and Sophistication from Procrastination and Preproperation." Discussion Paper dp16-11, Department of Economics, Simon Fraser University.

Galperti, Simone, and Bruno Strulovici. 2017. "A theory of intergenerational altruism." *Econometrica*, 85(4): 1175–1218.

Giné, Xavier, Dean Karlan, and Jonathan Zinman. 2010. "Put your money where your butt is: a commitment contract for smoking cessation." *American Economic Journal: Applied Economics*, 2(4): 213–235.

Gul, Faruk, and Wolfgang Pesendorfer. 2001. "Temptation and self-control." *Econometrica*, 69(6): 1403–1435.

Gul, Faruk, and Wolfgang Pesendorfer. 2004. "Self-control and the theory of consumption." *Econometrica*, 72(1): 119–158.

Gul, Faruk, and Wolfgang Pesendorfer. 2005. "The revealed preference theory of changing tastes." *Review of Economic Studies*, 72(2): 429–448.

Halevy, Yoram. 2015. "Time consistency: Stationarity and time invariance." *Econometrica*, 83(1): 335–352.

Hayashi, Takashi. 2003. "Quasi-stationary cardinal utility and present bias." *Journal of Economic Theory*, 112(2): 343–352.

Heidhues, Paul, and Botond Kőszegi. 2010. "Exploiting naïveté about self-control in the credit market." *American Economic Review*, 100(5): 2279–2303.

Heidhues, Paul, and Botond Kőszegi. 2017. "Naïveté-based discrimination." *Quarterly Journal of Economics*, 132(2): 1019–54.

Houthakker, Hendrik S. 1950. "Revealed preference and the utility function." *Economica*, 16(77): 159–174.

John, Anett. 2019. "When Commitment Fails – Evidence from a Regular Saver Product in the Philippines." *Management Science*, 66(2): 503–529.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan. 2010. "Self-control and the development of work arrangements." *American Economic Review P&P*, 100(2): 624–628.

Manzini, Paola, and Marco Mariotti. 2007. "Sequentially rationalizable choice." *American Economic Review*, 97(5): 1824–1839.

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y Ozbay. 2012. "Revealed Attention." *American Economic Review*, 102(5): 2183–2205.

Montiel Olea, José Luis, and Tomasz Strzalecki. 2014. "Axiomatization and measurement of quasi-hyperbolic discounting." *Quarterly Journal of Economics*, 129(3): 1449–1499.

Noor, Jawwad. 2011. "Temptation and revealed preference." *Econometrica*, 79(2): 601–644.

Noor, Jawwad, and Norio Takeoka. 2017. "Impatience as Selfishness." Working Paper.

O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing it now or later." *American Economic Review*, 89(1): 103–124.

O'Donoghue, Ted, and Matthew Rabin. 2001. "Choice and procrastination." *Quarterly Journal of Economics*, 116(1): 121–160.

Ok, Efe A. 2007. *Real Analysis with Economic Applications.* Princeton University Press.

Ok, Efe A, and Yusufcan Masatlioglu. 2007. "A theory of (relative) discounting." *Journal of Economic Theory*, 137(1): 214–245.

Pollak, Robert A. 1968. "Consistent planning." *Review of Economic Studies*, 35(2): 201–208.

Quah, John K-H, and Bruno Strulovici. 2013. "Discounting, values, and decisions." *Journal of Political Economy*, 121(5): 896–939.

Rubinstein, Ariel, and Yuval Salant. 2008. "Some thoughts on the principle of revealed preference." In *The Foundations of Positive and Normative Economics*. , ed. Andrew Caplin and Andrew Schotter, 116–124. Oxford University Press.

Sen, A.K. 1971. "Choice functions and revealed preference." *Review of Economic Studies*, 38(3): 307–317.

Sprenger, Charles. 2015. "Judging Experimental Evidence on Dynamic Inconsistency." *American Economic Review: Papers and Proceedings*, 105(5): 280–285.

Strotz, Robert. 1955. "Myopia and inconsistency in dynamic utility maximization." *Review of Economic Studies*, 23(3): 165–180.

Appendix: Proofs

Proof of Theorem 1

(i) $\implies$ (ii) Suppose c is observationally time consistent. Suppose $t_1 = c(A)$ and $t_3 = c(A \cup \{t_2\})$. If $t_3 \neq t_2$, then time consistency and $t_1 = c(A)$ implies that $t_1 = c(\{t_1, t_3\})$, but time consistency and $t_3 = c(A \cup \{t_2\})$ implies that $t_3 = c(\{t_1, t_3\})$, and thus, $t_3 = t_1$ must hold. This proves that c cannot exhibit a reversal.

(ii) $\implies$ (i)

Suppose that c exhibits no reversals. If $|A| = 2$, $t = c(A)$ and $t' \in A$, then $t = c(\{t, t'\})$ since $\{t, t'\} = A$. Thus observational time consistency holds for all $A \in \mathscr{A}$ with $|A| = 2$. Suppose that, for some $n > 2$, observational time consistency holds whenever $|A| < n$. Now consider some A with $|A| = n - 1$ and $t' \notin A$ and $t = c(A)$; let $A' = A \cup \{t'\}$. Let $t^c = c(A')$. By the no reversals property, $t^c = t$ or $t^c = t'$. For any $t'' \in (A') \backslash \{t^c\}$, the no reversals property implies that $c(A' \backslash \{t''\}) = c(A')$. By observational time consistency on $A' \backslash \{t''\}$, $c(A') = c(\{t^c, t'''\})$ for all $t''' \in A' \backslash \{t''\}$. But since the choice of t'' was arbitrary, $c(A \cup \{t'\}) = c(\{c(A \cup \{t'\}), t''\})$ as well. Thus observational time consistency is satisfied for $A \cup \{t'\}$.

(i) $\iff$ (iii)

First, show that c is observationally time consistent if and only if there exists a complete, transitive, and antisymmetric binary relation R such that $c(A)$ equals the R-maximal element in A for every $A \in \mathscr{A}$.

Define R by tRt' if and only if there exists a $A \in \mathscr{A}$ such that $t = c(A)$ and $t' \in A$. Since $c(\{t, t'\}) = t$ or $= t'$, R is complete. Since $t = c(A)$ and $t' \in A$ implies $t = c(\{t, t'\})$ while $t' = c(A')$ and $t \in A'$ implies $t' = c(\{t, t'\})$, R is antisymmetric. To show that R is transitive, suppose $t_1 R t_2$ and $t_2 R t_3$. Then by the definition of R and its antisymmetry, we must have (a) $t_1 = c(\{t_1, t_2\})$ and (b) $t_2 = c(\{t_2, t_3\})$. By (a) and observational time consistency, $t_2 \neq c(\{t_1, t_2, t_3\})$. Similarly by (b), $t_3 \neq c(\{t_1, t_2, t_3\})$. Thus $t_1 = c(\{t_1, t_2, t_3\})$, so $t_1 R t_3$, proving that R is symmetric.

Conversely, suppose that there exists a complete and transitive binary relation R such that $c(A)$ equals the R-maximal element in A for every $A \in \mathscr{A}$. Now take an arbitrary $A \in \mathscr{A}$ and suppose $t = c(A)$. Since t must be R-maximal in c, $tRt'$for all $t' \in A$, which implies $t = c(\{t,t'\})$ for all $t' \in A$. Thus c must be observationally time consistent.

Second, show that such a representation is equivalent to a Strotzian representation that is both sophisticated and naïve.

First, define a function $V : \bar{A} \to \mathbb{N}$ by $V(t) = |\{t' \in \bar{A} : tRt'\}|$. By transitivity and antisymmetry of R, tRt' implies $V(t) > V(t')$; conversely, by completeness and transitivity of R, $V(t) > V(t')$ implies tRt'. By antisymmetry of R, V is an injection. Thus $c(A) = \arg\max_{t \in A} V(t)$ for all A. For each t, define U_t as the restriction of V to the domain $A_{\geq t}$. For each t_1 and $t_2 > t_1$ define $\hat{U}_{t_2|t_1} = U_{t_2}$, and note that by construction, $\hat{U}_{t_2|t_1} = U_{t_1}$ on $\bar{A}_{\geq t_1}$. By choice of $\mathscr{U}$, working backward implies that at each t_1 and each $t_2 > t_1$ a person's perceived future decision $s(t_2, A, \hat{U}_{t_2|t_1}, \hat{\mathscr{U}}_{\cdot|t_1}) =$ act if and only if $t_2 = \arg\max_{t' \in A_{\geq t_2}} V(t')$, and similarly $s(t_1, A, U_{t_1}, \hat{\mathscr{U}}_{\cdot|t_1}) =$ act if and only if $t_1 = \arg\max_{t' \in A_{\geq t_1}} V(t')$. Thus $c(A) = \min\{t \in A : s(t, A, U_t, \hat{\mathscr{U}}_{\cdot|t}) = \text{act}\}$; by construction, $\mathscr{U}, \hat{\mathscr{U}}$ is a Strotzian representation that is both naïve and sophisticated.

□

Proof of Proposition 1

Suppose $t_1 = c(A)$, $t_3 = c(A \cup \{t_2\}) > t_1, t_2$. Then by the naïve representation, because $t_1 = c(A)$ it must be the case that $U_{t_1}(t_1) > U_{t_1}(t_3)$. Thus $t_1 = c(\{t_1, t_3\})$.

□

Proof of Theorem 4

Consider a reversal involving t_1, t_2, t_3 with $t_1 = c(A)$ and $t_3 = c(A \cup \{t_2\}) \neq t_1, t_2$.

Since c is naive, it cannot exhibit a doing-it-earlier reversal (Theorem 2) so $t_3 > t_1$.

Since $t_1 = c(A)$, by the naïve representation, $U_{t_1}(t_1) > U_{t_1}(t) \; \forall t \in A_{>t}$ (including for $t = t_3$); thus since $t_1 < c(A \cup \{t_2\})$, $U_{t_1}(t_2) > U_{t_1}(t_1)$. So $t_2 = c(\{t_1, t_2\})$, $t_1 = c(\{t_1, t_3\})$. By the naïve representation, $U_{\min\{t_2,t_3\}}(t_3) > U_{\min\{t_2,t_3\}}(t_2)$, thus $t_3 = c(\{t_2, t_3\})$. Thus, $t_2 = c(\{t_1, t_2\})$, $t_1 = c(\{t_1, t_3\})$, and $t_3 = c(\{t_2, t_3\})$.

(Proof of i) If $t_2 > t_3$, then $t_1 < t_3 < t_2$, and c cannot exhibit only procrastination-inducing inconsistency.

(Proof of ii) If instead $t_2 < t_3$, then $t_1 < t_2 < t_3$, and thus c cannot exhibit only preproperation-inducing inconsistency.

□

Proof of Theorem 5

Consider a reversal involving t_1, t_2, t_3 with $t_1 = c(A)$ and $t_3 = c(A \cup \{t_2\}) \neq t_1, t_2$.

(i) Since c is sophisticated, it must be the case that $t_2 = c(A_{>t_2} \cup \{t_2\})$; otherwise, by repeatly applying Recursivity, $c(A \cup \{t_2\}) = t_1$. Thus, it follows that $t_3 < t_2$.

(ii) Suppose $t_3 < t_2 < t_1$. Let $t_4 = \max\{t \in A_{<t_2} : t = c(A_{\geq t} \cup \{t_2\})\}$; by construction, $\{t \in A_{<t_2} : t = c(A_{\geq t} \cup \{t_2\})\}$ is finite and includes t_3, so t_4 is well-defined. Then by sophistication, $t_1 = c(\{t_4, t_1\})$, $c(\{t_2, t_1\}) = t_2$, $c(\{t_4, t_2\}) = t_4$. Thus c cannot exhibit only procrastination-inducing inconsistency.

(iii) Suppose $t_3 < t_1 < t_2$.

Let $t_4 = \max\{t \in A_{<t_1} : t = c(A_{\geq t} \cup \{t_2\})\}$ and let $t_5 = \min\{t \in A_{>t_1} : t = c(A_{\geq t} \cup \{t_2\})\}$. Since $t_3 \in \max\{t \in A_{<t_1} : t = c(A_{\geq t} \cup \{t_2\})\}$, $t_2 \in \{t \in A_{>t_1} : t = c(A_{\geq t} \cup \{t_2\})\}$ by sophistication, and both sets are finite, both maximizers exist. By the sophisticated representation, $t_1 = c(\{t_4, t_1\})$, $t_4 = c(\{t_4, t_5\})$, and $t_5 = c(\{t_1, t_5\})$. Since $t_4 < t_1 < t_5$, c cannot exhibit only preproperation-inducing inconsistency.

□

Proof of Theorem 6

Part (i): Necessity.

Suppose c has a naïve representation $\left(\mathscr{U}, \hat{\mathscr{U}}\right)$.

First, suppose that $t_1 = c(A)$ and $t_3 = c(A \cup \{t_2\})$ where $t_1 \neq t_3 \neq t_2$. Since for each t, $\max\limits_{t' \in (A \cup \{t_2\})_{>t}} U_t(t') \geq \max\limits_{t' \in A_{>t}} U_t(t')$ and by naïveté $\hat{U}_{t'|t}(t'') = U_t(t'')$ for any

$t < t' \leq t''$, we have that $s(t, A, U_t, \hat{\mathscr{U}}_{\cdot|t}) =$ wait implies $s(t, A \cup \{t_2\}, U_t, \hat{\mathscr{U}}_{\cdot|t}) =$ wait for each $t \neq t_2$. Since $t_2 \neq t_3$, it follows from the naïve representation that $t_3 > t_1$.

But since $s(t_1, A, U_{t_1}, \hat{\mathscr{U}}_{\cdot|t_1}) =$ act and $s(t_1, A \cup \{t_2\}, U_{t_1}, \hat{\mathscr{U}}_{\cdot|t_1}) =$ wait, it follows by the representation that $\max_{t' \in A_{>t_1}} U_{t_1}(t') < U_{t_1}(t_1) < \max_{t' \in (A \cup \{t_2\})_{>t_1}} U_{t_1}(t')$, thus it must be the case that $t_2 > t_1$ and $U_{t_1}(t_2) > U_{t_1}(t_1)$. Since $t_2, t_3 > t_1$, this proves that c satisfies the Irrelevant Alternatives Delay property. In addition, since $U_{t_1}(t_2) > U_{t_1}(t_1)$, for any A' with $t_2 \in A'$, $U_{t_1}(t_1) < U_{t_1}(t_2) \leq \max_{t' \in A'_{>t_1}} U_{t_1}(t')$, thus wait $= s(t_1, A', U_{t_1}, \hat{\mathscr{U}}_{\cdot|t_1})$, thus $t_1 \neq c(A')$. This proves that c satisfies Exclusion Consistency.

Part (i): Sufficiency.

Consider the following, alternative expression of the definition of naïve perception-perfect equilibrium:

(i) $A_{>t} = \emptyset$ implies $s(t, A, U_t, \hat{\mathscr{U}}_{\cdot|t}) =$ act,

(ii) $U_t(t) > \max_{t' \in A_{>t}} U_t(t')$ implies $s(t, A, U_t, \hat{\mathscr{U}}_{\cdot|t}) =$ act, and

(iii) $U_t(t) < \max_{t' \in A_{>t}} U_t(t')$ implies $s(t, A, U_t, \hat{\mathscr{U}}_{\cdot|t}) =$ wait.

This definition gives $t = c(A)$ if and only if it is the case that $U_t(t) > U_t(t')$ for all $t' \in A_{\geq t}$, whereas for each $t' \in A_{<t}$, there exists a $t'' \in A_{\geq t'}$ such that $U_{t'}(t'') > U_{t'}(t')$.

Construct utility functions from choices. Construct U_t by setting $U_t(t) > U_t(t')$ if $t' > t$ and $t = c(\{t, t'\})$, and setting $U_t(t') > U_t(t)$ if $t' = c(\{t, t'\})$.

Such a U_t is well-defined, since if $t_1 < t_2, t_3$ and c has the cycle with (i) $t_1 = c(\{t_1, t_2\})$, (ii) $t_2 = c(\{t_2, t_3\})$ and $t_3 = c(\{t_1, t_3\})$, we would conclude that $U_{t_1}(t_3) > U_{t_1}(t_1) > U_{t_1}(t_2)$ and $U_{\min\{t_2, t_3\}}(t_2) > U_{\min\{t_2, t_3\}}(t_3)$. Intuitively, in this construction of U_{t_1}, choice only dictates for each $t_2 > t_1$ whether $U_{t_1}(t_1) \gtrless U_{t_1}(t_2)$; thus choice cannot generate cycles that prevent such a construction. Thus U_t is well-defined.

Show that the naïve perception-perfect equilibrium option in a set is chosen. Suppose $t_1 < \cdots < t_n$ and for some $j > i$ we have $U_{t_i}(t_i) < U_{t_i}(t_j)$. Then by the definition of U_{t_i}, $t_j = c(\{t_i, t_j\})$. Irrelevant Alternatives Delay requires that a newly added option that leads to a reversal must lead to delay. Since and $t_i < t_k$ for all $k > i$, by repeatedly applying Irrelevant Alternatives Delay we have:

$$\begin{aligned} t_i &\neq c(\{t_i, t_{i+1}, t_j\}) \\ &\vdots \\ t_i &\neq c(\{t_i, \dots, t_n\}) \end{aligned}$$

Irrelevant Alternatives Delay requires that if a newly added option leads to a reversal, it must be a later option than that originally chosen. Since $k < k'$ implies $t_k < t_{k'}$, by repeatedly applying Irrelevant Alternatives Delay we have:

$$\begin{aligned} t_i &\neq c(\{t_{i-1}, t_i, \dots, t_n\}) \\ &\vdots \\ t_i &\neq c(\{t_1, \dots, t_n\}) \end{aligned}$$

Thus, if $t_j > t_i$, $U_{t_i}(t_j) > U_{t_i}(t_i)$, and $t_j \in A$, then $t_i \neq c(A)$. Call this result (A).

Next, suppose that $t_1 < \cdots < t_n$ and $U_{t_1}(t_1) > U_{t_1}(t_i)$ for all $i > 1$. Then by the definition of U_{t_1}, $t_1 = c(\{t_1, t_i\})$ for each i. Thus, by Exclusion Consistency, for each i and j, we have $t_1 = c(\{t_1, t_i, t_j\})$. Now suppose that for any $A \subseteq \{t_1, \dots t_n\}$ with $|A| < k$, we have $t_1 = c(A)$. Now consider A with $|A| = k$. Since $t_1 = c(A \setminus \{t_i\})$ for each $i \neq 1$, and $t_1 = c(\{t_1, t_j\})$, by Exclusion Consistency it must be the case that $t_1 = c(A)$ (since picking $i \neq j$ would generate a contradiction if $t_j = c(A)$ for $j > 1$). By applying this argument until $k = n$, we obtain that if $t_1 < \cdots < t_n$ and $U_{t_1}(t_1) > U_{t_1}(t_i)$ for all $i > 1$, then $t_1 = c(\{t_1, \dots, t_n\})$. Call this result (B).

Next, suppose $t_1 < \cdots < t_{i-1} < t_i < \cdots < t_n$. Suppose further that $U_{t_i}(t_i) > U_{t_i}(t_j)$ for all $j > i$, and for each $j < i$, there exists a $k > j$ such that $U_{t_j}(t_k) > U_{t_j}(t_j)$. Then, by the definition of U_{t_i}, $t_i = c(\{t_i, t_j\})$ for each $j > i$. Then by result (B), we have $t_i = c(\{t_i, \dots, t_n\})$. So now suppose that for some $j \leq i$, we have that $t_i = c(\{t_j, \dots, t_n\})$. Then by result (A), $t_{j-1} \neq c(\{t_{j-1}, \dots, t_n\})$; but since $t_{j-1} < t_i$, by Irrelevant Alternatives Delay, it follows that $t_i = c(\{t_{j-1}, \dots, t_n\})$.

Repeating the argument until $j=1$ yields that $t_i = c(\{t_1,\ldots,t_n\})$. Thus, if t_i is the naïve perception-perfect equilibrium prediction for $\{t_1,\ldots,t_n\}$ under $\mathscr{U}$, then $t_i = c(\{t_1,\ldots,t_n\})$. Call this result (C).

Show that the chosen option is the naïve perception-perfect equilibrium. Suppose $t_i = c(\{t_1,\ldots,,t_n\})$. Given $\mathscr{U}$, the naïve perception-perfect equilibrium concept makes a unique choice prediction; by result (C), this must be equal to t_i.

Part (ii): Necessity.

Suppose c has a sophisticated representation $\left(\mathscr{U},\hat{\mathscr{U}}\right)$ and suppose $t' = c(A_{>t})$ and $t \in A$.

Then by the sophisticated representation,

$$\begin{aligned} t' &= \min\left\{\tau : s(\tau, A_{>t}, U_\tau, \hat{\mathscr{U}}_{\cdot|\tau}) = \text{act}\right\} \\ &= \min\left\{\tau > t : s(\tau, A_{\geq t}, U_\tau, \hat{\mathscr{U}}_{\cdot|\tau}) = \text{act}\right\} \\ &= \min\left\{\tau > t : s(\tau, A_{\geq t}, U_\tau, \hat{\mathscr{U}}_{\cdot|t}) = \text{act}\right\} \end{aligned}$$

Thus,

$$s(t, A_{\geq t}, U_t, \hat{\mathscr{U}}_{\cdot|t}) = \begin{cases} \text{act} & \text{if } U_t(t) > U_t(t') \\ \text{wait} & \text{if } U_t(t) < U_t(t') \end{cases}.$$

Thus, if $U_t(t) > U_t(t')$, then $t = \min\left\{\tau : s(\tau, A_{\geq t}, U_t, \hat{\mathscr{U}}_{\cdot|t}) = \text{act}\right\} = c(A_{\geq t})$. If instead $U_t(t) < U_t(t')$, then $t' = \min\left\{\tau : s(\tau, A_{\geq t}, U_t, \hat{\mathscr{U}}_{\cdot|t}) = \text{act}\right\} = c(A_{\geq t})$. Since these cases are exhaustive, it follows that Recursivity holds.

Part (ii): Sufficiency.

Construct $\{U_t\}_{t=1}^{T}$ using choices from two-element sets as in the proof in part (i) – the same arguments directly apply. For each t and each $t' > t$, define $\hat{U}_{t'|t} = U_{t'}$ on $\bar{A}_{\geq t'}$. It follows by this construction that $c(\{t_1,t_2\}) = \min\{t \in \{t_1,t_2\} : s(t, \{t_1,t_2\}, U_t, \hat{\mathscr{U}}_{\cdot|t}) = \text{act}\}$ for all $t_1, t_2 \in \bar{A}$, establishing a sophisticated representation on two-element sets. Now proceed by induction.

Suppose that $c(A) = \min\{t \in A : s(t, A, \mathscr{U}, \hat{\mathscr{U}}_{\cdot|t}) = t\}$ and $t_0 < \min A$. Applying Recursivity and then the construction of U_{t_0},

$$
\begin{aligned}
c(A \cup \{t_0\}) &= c(\{t_0, c(A)\}) \\
&= \begin{cases} t_0 & \text{if } U_{t_0}(t_0) > U_{t_0}(c(A)) \\ c(A) & \text{otherwise} \end{cases}
\end{aligned}
$$

Thus, since $s(t_0|A \cup \{t_0\}, \mathscr{U}, \hat{\mathscr{U}}^{s|\mathscr{U}}) = \begin{cases} t_0 & \text{if } U_{t_0}(t_0) > U_{t_0}(\hat{\tau}_{t_0}) \\ \text{wait} & \text{otherwise} \end{cases}$ for $\hat{\tau}_{t_0} = \min\{t \in A : s(t, A \cup \{t_0\}, U_t, \hat{\mathscr{U}}_{\cdot|t}) = t\} = \min\{t \in A : s(t, A, U_t, \hat{\mathscr{U}}_{\cdot|t}) = t\} = c(A)$, and $s(t, A \cup \{t_0\}, U_t, \hat{\mathscr{U}}_{\cdot|t}) = s(t, A, U_t, \hat{\mathscr{U}}_{\cdot|t})$ for all $t > t_0$, it follows that $c(A \cup \{t_0\}) = \min\{t \in A \cup \{t_0\} : s(t, A \cup \{t_0\}, U_t, \hat{\mathscr{U}}_{\cdot|t}) = \text{act}\}$. The sophisticated representation thus extends to all of $\mathscr{A}$ by induction.

□

Proposition 5. *If c satisfies Penultimate Replaceability, then c satisfies End Replaceability.*

Proof of Proposition 5

Let c satisfy Penultimate Replaceability. Fix $t_1 < t_2 < \cdots < t_n$. Given the set $\{t_2, \ldots t_{k-1}, \tilde{t}^k\}$, Penultimate Replaceability implies that there exists $\tilde{t}^{k-1} \in \{t_{k-1}, \tilde{t}^k\}$ such that for all A with $t_1 \in A = A_{<t_2}$, $c_{t_1}(A \cup A^k) = c_{t_1}(A \cup \{t_2, \ldots t_{k-2}, \tilde{t}^{k-1}\})$. Let $A' = A^n$ and start with $k = n$. Given any $k > 2$, apply this argument to obtain $\tilde{t}^{k-1} \in \{t_{n-1}, \tilde{t}^k\}$ such that $c_{t_1}(A \cup A^k) = c_{t_1}(A \cup \{t_2, \ldots t_{k-2}, \tilde{t}^{k-1}\})$ and set $A^{k-1} = \{t_2, \ldots, t_{k-2}, \tilde{t}^{k-1}\}$. Since $\tilde{t}^n \in \{t_{n-1}, t_n\}$, and $\tilde{t}^{k-1} \in \{t_{k-2}, \tilde{t}^k\}$ for each $k < n$, combining these restrictions give $\tilde{t}^2 \in A'$. Thus, there exists $t_k \in A'$, specifically $t_k = \tilde{t}^2$, such that $c_{t_1}(A \cup \{t_k\}) = c_{t_1}(A \cup A')$ for all A such that $t_1 \in A = A_{<t_2}$.

□

Proof of Theorem 7

Necessity.

(i) Suppose c has a Strotzian representation $\left(\mathscr{U},\hat{\mathscr{U}}\right)$. Then, given any t_0,t_1,t_2 with $t_0<t_1<t_2$ we either have $\hat{U}_{t_1|t_0}(t_1)>\hat{U}_{t_1|t_0}(t_2)$ or $\hat{U}_{t_1|t_0}(t_2)>\hat{U}_{t_1|t_0}(t_1)$. In the former case, for all $A\subseteq\{t_0,\dots,t_1-1\}$ we have $t_1=s(t_1,A\cup\{t_1,t_2\},\hat{U}_{t_1|t_0},\hat{\mathscr{U}}_{\cdot|t_0})$, and thus $s(t,A\cup\{t_1,t_2\},\hat{U}_{t|t_0},\hat{\mathscr{U}}_{\cdot|t_0})=s(t,A\cup\{t_1\},\hat{U}_{t|t_0},\hat{\mathscr{U}}_{\cdot|t_0})$ for all t with $t_0<t<t_1$, thus, $c_{t_0}(A\cup\{t_1,t_2\})=s(t_0,A\cup\{t_1,t_2\},U_{t_0},\hat{\mathscr{U}}_{\cdot|t_0})=s(t_0,A\cup\{t_1\},U_{t_0},\hat{\mathscr{U}}_{\cdot|t_0})$. An analogous argument applies in the latter case, implying $c_{t_0}(A\cup\{t_1,t_2\})=s(t_0,A\cup\{t_1,t_2\},U_{t_0},\hat{\mathscr{U}}_{\cdot|t_0})=s(t_0,A\cup\{t_2\},U_{t_0},\hat{\mathscr{U}}_{\cdot|t_0})$ for all $A\subseteq\{t_0,\dots,t_1-1\}$. Thus, c satisfies Penultimate Replaceability.

(ii) Next, suppose that c has partially naïve representation $\left(\mathscr{U},\hat{\mathscr{U}}\right)$, and $c(\{t_1,t_2\})=c(\{t_0,t_2\})=t_2$ but $c(\{t_0,t_1\})=t_0$. Thus, $U_{\min\{t_1,t_2\}}(t_2)>U_{\min\{t_1,t_2\}}(t_1)$ and $U_{t_0}(t_2)>U_{t_0}(t_0)>U_{t_0}(t_1)$. Then by partially naïve restriction, $\hat{U}_{\min\{t_1,t_2\}|t_0}(t_2)>\hat{U}_{\min\{t_1,t_2\}|t_0}(t_1)$. By the Strotzian representation, $t_2=c(\{t_0,t_1,t_2\})\neq t_0$. Thus, c must satisfy Wishfulness.

Sufficiency.

(i) Let c satisfy Penultimate Replaceability.

Construct $\left(\mathscr{U},\hat{\mathscr{U}}\right)$ to satisfy requirements 1 and 2 below.

Requirement 1. If $t_1=c(\{t_1,t_2\})$, then require $U_{\min\{t_1,t_2\}}(t_1)>U_{\min\{t_1,t_2\}}(t_2)$.

Requirement 2. If $t_1,t_2>t_0$, and $\exists A$ such that $t_0\in A\subseteq\{t_0,\dots,\min\{t_1,t_2\}-1\}$ and $c_{t_0}(A\cup\{t_1,t_2\})=c_{t_0}(A\cup\{t_1\})\neq c_{t_0}(A\cup\{t_2\})$, then require that $\hat{U}_{\min\{t_1,t_2\}|t_0}(t_1)>\hat{U}_{\min\{t_1,t_2\}|t_0}(t_2)$. By Penultimate Replaceability, each $\hat{U}_{\min\{t_1,t_2\}|t_0}$ can be constructed to satisfy this requirement.

Given $\hat{\mathscr{U}}$, define $\hat{c}^{|t_0}$ mapping from subsets of $\mathscr{A}_{>t_0}$ to $\bar{A}_{>t_0}$, by $\hat{c}^{|t_0}(A_{>t_0})=\min\{t\in A_{>t_0}:s(t,A,\hat{U}_{t|t_0},\hat{\mathscr{U}}_{\cdot|t_0})=\text{act}\}$. The function $\hat{c}^{|t_0}$ is the time t_0 self's expected future choice function in a Strotzian representation.

It remains to show that for arbitrary t_0 and A that $c_{t_0}(\{t_0\}\cup A_{>t_0})=c_{t_0}(\{t_0,\hat{c}^{|t_0}(A_{>t_0})\})$. We know that the representation holds whenever $|A_{>t_0}|=1$ by the representation on two element choice sets. So suppose that, for some $m\geq 2$, $c_{t_0}(\{t_0,\dots,t_n\})=c_{t_0}(\{t_0,\hat{c}^{|t_0}(\{t_1,\dots,t_n\})\})$ whenever $n<m$ (where $t_0<t_1<\cdots<t_n$). Now let $n=m$ and consider $\{t_0,\dots,t_n\}$. By Penultimate Replaceability,

$c_{t_0}(\{t_0,\ldots,t_n\}) = c_{t_0}(\{t_0,\ldots,t_{n-2},t_n\})$ or $= c_{t_0}(\{t_0,\ldots,t_{n-1}\})$; by the construction of $\hat{U}_{\min\{t_1,t_2\}|t_0}$ and $\hat{c}^{|t_0}$, we have $c_{t_0}(\{t_0,\ldots,t_n\}) = c_{t_0}(\{t_0,\ldots,t_{n-2},\hat{c}^{|t_0}(\{t_{n-1},t_n\})\})$. Since $\{t_1,\ldots,t_{n-2},\hat{c}^{|t_0}(\{t_{n-1},t_n\})\}$ has only $m-1$ elements, it follows that

$$c_{t_0}(\{t_0,\ldots,t_n\}) = c_{t_0}(\{t_0,\hat{c}^{|t_0}(\{t_1,\ldots,t_{n-2},\hat{c}^{|t_0}(\{t_{n-1},t_n\})\})\}) \quad (1)$$

Applying our recursive definition of $\hat{c}^{|t_0}$, working forward from t_1,

$$\begin{aligned}
\hat{c}^{|t_0}(\{t_1,\ldots,t_{n-1},\hat{c}^{|t_0}(\{t_{n-1},t_n\})\}) &= \hat{c}^{|t_0}(\{t_1,\hat{c}^{|t_0}(\{t_2,\ldots,t_{n-2},\hat{c}^{|t_0}(\{t_{n-1},t_n\})\})\}) \\
&= \hat{c}^{|t_0}(\{t_1,\hat{c}^{|t_0}(\{t_2,\hat{c}^{|t_0}(\{t_3,\ldots,t_{n-2},\hat{c}^{|t_0}(\{t_{n-1},t_n\})\})\})\}) \\
\vdots \quad &\vdots \quad \vdots \\
&= \hat{c}^{|t_0}(\{t_1,\hat{c}^{|t_0}(\{t_2,\hat{c}^{|t_0}(\{t_3,\hat{c}^{|t_0}(\{t_3,\ldots\})\})\})\}) \\
&= \hat{c}^{|t_0}(\{t_1,\ldots,t_n\}) \quad (2)
\end{aligned}$$

Then, working backward from the final expression, repeatedly apply the definition of $\hat{c}^{|t_0}$ to simplify: $\hat{c}^{|t_0}(\{t_{n-2},\hat{c}^{|t_0}(\{t_{n-1},t_n\})\}) = \hat{c}^{|t_0}(\{t_{n-2},t_{n-1},t_n\})$, $\hat{c}^{|t_0}(\{t_{n-3},\hat{c}^{|t_0}(\{t_{n-2},t_{n-1},t_n\})\}) = \hat{c}^{|t_0}(\{t_{n-3},t_{n-2},t_{n-1},t_n\})$, and so on, until obtaining $\hat{c}^{|t_0}(\{t_1,\hat{c}^{|t_0}(\{t_2,\ldots,t_n\})\}) = \hat{c}^{|t_0}(\{t_1,\ldots,t_n\})$. Combining this with (1) and (2), obtain that $c_{t_0}(\{t_0,\ldots,t_n\}) = c_{t_0}(\{t_0,\hat{c}^{|t_0}(\{t_1,\ldots,t_n\})\})$. Repeating the preceding steps of argument establishes this result, and thus the Strotzian representation for arbitrary choice sets of size n; the representation holds for arbitrary sets by induction. This proves part (i).

Next, suppose that c also satisfies Wishfulness.

For each $t \in \bar{A}$, R_t is acyclic, each has an asymmetric transitive closure, denote it by $\bar{R}_t$. By Szpilrajn's Theorem (see Ok 2007, p. 17) each $\bar{R}_t$ has a transitive and asymmetric completion $\bar{\bar{R}}_t$, which hence has a one-to-one utility representation U_t. Based on this reasoning, consider the following strengthening of Requirement 1 to 1' and the additional Requirement 3.

Requirement 1'. For each $t < T$, U_t represents a transitive and asymmetric completion $\bar{\bar{R}}_t$ of R_t.

Requirement 3. If $t_1,t_2 > t_0$ and $\forall A$ such that $t_0 \in A \subseteq \{t_0,\ldots,\min\{t_1,t_2\} -$

$1\}$, $c_{t_0}(A \cup \{t_1,t_2\}) = c_{t_0}(A \cup \{t_1\}) = c_{t_0}(A \cup \{t_2\})$, $t_1 = c(\{t_1,t_2\})$, then require $\hat{U}_{\min\{t_1,t_2\}|t_0}(t_1) > \hat{U}_{\min\{t_1,t_2\}|t_0}(t_2)$.

By Penultimate Replaceability and Wishfulness, $\mathscr{U}$ and $\hat{\mathscr{U}}$ can be constructed to satisfy Requirements 1', 2, and 3.

It remains to show the partial naïve restriction holds. Take any t_1. From the definition of R_{t_1} and construction of $U_{\min\{t_2,t_3\}}$, $U_{\min\{t_2,t_3\}}(t_2) > U_{\min\{t_2,t_3\}}(t_3)$ and $\hat{U}_{\min\{t_2,t_3\}|t_1}(t_3) > \hat{U}_{\min\{t_2,t_3\}|t_1}(t_2)$ implies that $t_3 R_{t_1} t_2$, which implies $U_{t_1}(t_3) > U_{t_1}(t_2)$ – thus implying that the representation is partially naïve. Further, this constructed representation is minimally naïve.

□

Proof of Corollary 1

Take c and c' with minimally naïve representations with the same $\mathscr{U}$ and different $\hat{\mathscr{U}}$ and $\hat{\mathscr{U}}'$.

Suppose $\left(\mathscr{U}, \hat{\mathscr{U}}\right)$ is more sophisticated than $\left(\mathscr{U}, \hat{\mathscr{U}}'\right)$. If $t_1,t_2 > t_0$ and $\forall A$ such that $t_0 \in A \subseteq \{t_0,\ldots,\min\{t_1,t_2\}-1\}$, $c'_{t_0}(A \cup \{t_1,t_2\}) = c'_{t_0}(A \cup \{t_1\}) \neq c'_{t_0}(A \cup \{t_2\})$ and $t_1 = c'(\{t_1,t_2\})$, then $\hat{U}'_{\min\{t_1,t_2\}|t_0}(t_1) > \hat{U}'_{\min\{t_1,t_2\}|t_0}(t_2)$ and $U_{\min\{t_1,t_2\}}(t_1) > U_{\min\{t_1,t_2\}}(t_2)$ by the representation for c'. Since $\mathscr{U}, \hat{\mathscr{U}}$ is more sophisticated than $\mathscr{U}, \hat{\mathscr{U}}'$, it follows that $\hat{U}_{\min\{t_1,t_2\}|t_0}(t_1) > \hat{U}_{\min\{t_1,t_2\}|t_0}(t_2)$, and thus by the representation for c, that $c_{t_0}(A \cup \{t_1,t_2\}) = c_{t_0}(A \cup \{t_1\}) = c_{t_0}(A \cup \{c(\{t_1,t_2\})\})$. Thus c is more penultimately accurate than c'.

Next, suppose c is more penultimately accurate than c'.

Suppose $\exists A$ such that $t_0 \in A \subseteq \{t_0,\ldots,\min\{t_1,t_2\}-1\}$, $c'_{t_0}(A \cup \{t_1,t_2\}) = c'_{t_0}(A \cup \{t_1\}) \neq c'_{t_0}(A \cup \{t_2\})$, then $\hat{U}'_{\min\{t_1,t_2\}|t_0}(t_1) > \hat{U}'_{\min\{t_1,t_2\}|t_0}(t_2)$ must hold by the representation. Consider two subcases.

First, if $t_1 = c'(\{t_1,t_2\})$, then $U_{\min\{t_1,t_2\}}(t_1) > U_{\min\{t_1,t_2\}}(t_2)$ by the representation for c'. Thus $t_1 = c(\{t_1,t_2\})$ by the representation for c, and thus since c is more penultimately accurate than c', $c_{t_0}(A \cup \{t_1,t_2\}) = c_{t_0}(A \cup \{t_1\})$ for all $A \subseteq \{t_0,\ldots,\min\{t_1,t_2\}-1\}$. Therefore, a minimally naïve representation for c must have $\hat{U}_{\min\{t_1,t_2\}|t_0}(t_1) > \hat{U}_{\min\{t_1,t_2\}|t_0}(t_2)$. In this subcase, $\hat{U}_{\min\{t_1,t_2\}|t_0}$ must be equally sophisticated as $\hat{U}'_{\min\{t_1,t_2\}|t_0}$ about that comparison.

Second, if $t_2 = c'(\{t_1, t_2\})$, then $U_{t_0}(t_1) > U_{t_0}(t_2)$ since $\mathscr{U}, \hat{\mathscr{U}}'$ is a partially naïve representation for c'. Since $\hat{U}'_{\min\{t_1,t_2\}|t_0}$ is naïve about this comparison in this case, the definition of more sophisticated than puts no restriction on $\hat{U}_{\min\{t_1,t_2\}|t_0}$ about this comparison so $\hat{U}_{\min\{t_1,t_2\}|t_0}$ is (trivially) more sophisticated than $\hat{U}'_{\min\{t_1,t_2\}|t_0}$ about this comparison.

Since the choice of t_0, t_1, t_2 was arbitrary, it follows that $\left(\mathscr{U}, \hat{\mathscr{U}}\right)$ is more sophisticated than $\left(\mathscr{U}, \hat{\mathscr{U}}'\right)$.

□

Proof of Proposition 2

(i) $\implies$ (ii). Suppose condition (i) holds, and c' exhibits doing-it-later reversal $c'(\{t_1, t_2, t_3\}) = t_3 > t_1 = c(\{t_1, t_3\})$. Then, $U_{t_1}(t_1) > U_{t_1}(t_3)$, thus by the Strotzian representation, waiting at t_1 when facing $\{t_1, t_2, t_3\}$ implies that $U_{t_1}(t_2) > U_{t_1}(t_1)$ and $\hat{U}'_{\min\{t_2,t_3\}|t_1}(t_2) > \hat{U}'_{\min\{t_2,t_3\}|t_1}(t_3)$; the fact that the decision maker subsequently acts at t_3 implies $U_{\min\{t_2,t_3\}}(t_3) > U_{\min\{t_2,t_3\}}(t_2)$. Then by (i), $\hat{U}_{\min\{t_2,t_3\}|t_1}(t_2) > \hat{U}_{\min\{t_2,t_3\}|t_1}(t_3)$, and it follows from the Strotzian representation for c that $c(\{t_1, t_2, t_3\}) = t_3 > t_1 = c(\{t_1, t_3\})$. Thus, (i) implies (ii).

(ii) $\implies$ (iii). Now suppose that (ii) holds. Suppose that $t_1 < t_2, t_3$ generates the doing-it-earlier reversal $c(\{t_1, t_2, t_3\}) = t_1 < t_3 = c(\{t_1, t_3\})$ for c (if no such t_1, t_2, t_3 exist, then c exhibits no three opportunity doing-it-earlier reversals, so the desired conclusion holds trivially). Since $t_3 = c(\{t_1, t_3\})$, the Strotzian representation implies that $U_{t_1}(t_3) > U_{t_1}(t_1)$. Since $t_1 = c(\{t_1, t_2, t_3\})$, it must also be the case that $U_{t_1}(t_1) > U_{t_1}(t_2)$ and $\hat{U}_{\min\{t_2,t_3\}|t_1}(t_2) > \hat{U}_{\min\{t_2,t_3\}|t_1}(t_3)$. Since c has a partially naïve representation, it follows that $U_{\min\{t_2,t_3\}}(t_2) > U_{\min\{t_2,t_3\}}(t_3)$. If $\hat{U}'_{\min\{t_2,t_3\}|t_1}(t_2) < \hat{U}'_{\min\{t_2,t_3\}|t_1}(t_3)$, then by the representation for c', $t_2 = c'(\{t_1, t_2, t_3\})$; but by the representation for c' it is also the case that $t_2 > t_1 = c'(\{t_1, t_2\})$, thus this is a doing-it-later reversal for c' not exhibited by c, which contradicts that c exhibits more three opportunity doing-it-later reversals than c'. Thus, $\hat{U}'_{\min\{t_2,t_3\}|t_1}(t_2) > \hat{U}'_{\min\{t_2,t_3\}|t_1}(t_3)$ and $t_1 = c'(\{t_1, t_2, t_3\})$ and $t_3 = c'(\{t_1, t_3\})$, proving (iii).

(iii) $\implies$ (i). Now suppose that (iii) holds, $\hat{U}_{\min\{t_2,t_3\}|t_1}(t_2) \gtrless \hat{U}_{\min\{t_2,t_3\}|t_1}(t_3)$ is three opportunity revealable, $U_{\min\{t_2,t_3\}}(t_2) > U_{\min\{t_2,t_3\}}(t_3)$, and $\hat{U}_{\min\{t_2,t_3\}|t_1}(t_2) >$

$\hat{U}_{\min\{t_2,t_3\}|t_1}(t_3)$. Then since c has a partially naïve representation and $\hat{U}_{\min\{t_2,t_3\}|t_1}(t_2) > \hat{U}_{\min\{t_2,t_3\}|t_1}(t_3)$ is three opportunity revealable, it must be the case that either (a) $U_{t_1}(t_3) > U_{t_1}(t_1) > U_{t_1}(t_2)$ or (b) $U_{t_1}(t_2) > U_{t_1}(t_1) > U_{t_1}(t_3)$. In case (b), since c' has a partially naïve representation and $U_{\min\{t_2,t_3\}}(t_2) > U_{\min\{t_2,t_3\}}(t_3)$, it follows that $\hat{U}'_{\min\{t_2,t_3\}|t_1}(t_2) > \hat{U}'_{\min\{t_2,t_3\}|t_1}(t_3)$. In case (a), the representation for c implies that $c(\{t_1,t_2,t_3\}) = t_1 < t_3 = c(\{t_1,t_3\})$, a doing-it-earlier reversal. But since c' exhibits more doing-it-earlier reversals than c, $c'(\{t_1,t_2,t_3\}) = t_1 < t_3 = c'(\{t_1,t_3\})$ as well. This implies that $\hat{U}'_{\min\{t_2,t_3\}|t_1}(t_2) > \hat{U}'_{\min\{t_2,t_3\}|t_1}(t_3)$ – the desired result.

□