Evolving Preferences and Naivete (ECON0106 - Written Assignment)

[redacted]

The evolution of preferences over time poses many challenges to decision makers (DMs). As explained by Mihm and Ozbek (2019), while policies are enacted today, they will only be implemented in the future and, therefore, their impact will depend on future rather than current preferences of affected agents. Similarly, agents victim of present bias might be bearing excessive costs of a gym subscription, by over-estimating their future attendance. Given these dynamic inconsistencies, to avoid taking inter-temporally sub-optimal decisions, it becomes crucial to understand whether and to what extent DMs anticipate how preferences evolve.

In this work, we review four articles to illustrate how recent theoretical literature tackles these questions. Anticipation of future preferences is conceptualised through the notion of sophistication and naivete of agents. Noor (2011) is among the first to provide a behavioural definition of sophistication, that is, correctly anticipating future tastes. Sophistication plays a key role in explaining a desire for commitment in a model in which preferences evolve over time and choices are driven by a struggle between normative judgments and temptation. Ahn et al. (2019) and Tang and Zhang (2023) provide behavioural definitions of naivete (which, opposed to sophistication, is the inability to anticipate future tastes) and characterise the implied utility representations, thus making the concept tractable for applications.

Thanks to these behavioural definitions of absolute and comparative naivete, it is possible to test whether and to what extent agents anticipate the evolution of preferences by comparing *ex ante* intended and *ex post* realised behaviour. Mihm and Ozbek (2019) describe a setting in which it is possible to elicit future preferences starting from current ones and making use of an objective (yet incomplete) ranking that is not evolving over time. This contribution provides a benchmark, against which we can compare actual behaviour of naive agents. In sum, the papers reviewed in this work establish a framework to test and characterise naive behaviour.

Anticipating future preferences

Mihm and Ozbek (2019) present a finite horizon decision framework in which the DM can anticipate future preference from current ones by making use of a dominance relation. X_t denotes the set of outcomes in each period $t = 0, \dots, T$, while $\mathcal{K}(X_t)$ is the collection of all non-empty finite subsets of X_t, that is, all possible menus at time t. Decision problems are defined recursively as in Gul and Pesendorfer (2005): $\mathcal{X}_T = \mathcal{K}(X_T)$ is the set of time T decision problems, while $\mathcal{X}_t = \mathcal{K}(X_t \times \mathcal{X}_{t+1})$ is the set of time $t = 1, \dots, T-1$ decision problems. The set H_t contains t-periods histories $h = (x_0, \dots, x_t)$ (sequences of outcomes), and $H := \cup_{t=0}^{T-1} H_t$ is the set of all possible histories. Given a history h, the DM ranks alternatives through a history dependent preference relation $\succsim_h$ defined on $X_{|h|} \times \mathcal{X}_{|h|+1}$. The collection of history dependent preferences forms a ranking system $(\succsim_H) := (\succsim_h)_{h \in H}$, which essentially captures how preferences evolve over time. Furthermore, Mihm and Ozbek (2019) introduce the notion of a plan φ, that is a mapping between decision problems from one period to the following one.[1]

The primitive in Mihm and Ozbek (2019) is a weak order $\succsim$ on $X_0 \times \mathcal{X}_1$, which describes preferences over initial decision problems. Such preference relation can be interpreted as an *ex ante* preference before the unfolding of histories. $\succsim$ is defined to be consistent when (i) decision problems at time 0 are compatible (analogous) with (the iterations of) a plan φ and (ii) such plan φ is rationalised by the ranking system $(\succsim_H)$, meaning that at each history h, decisions are taken optimally according to all $\succsim_h \in (\succsim_H)$. Abusing notation (but following Mihm and Ozbek (2019) in this), we say that $\succsim$ ranks alternatives at history h, where alternatives are denoted by (h, x, a) with $x \in X_{|h|}$ and $a \in \mathcal{X}_{|h|+1}$. This allows to directly compare the anticipated ranking $\succsim$ at history h with the realised ranking $\succsim_h$.

[1]Technically, Mihm and Ozbek (2019) introduce the notion of decision nodes and formally define φ as a map between decision nodes. Such notation is avoided here to keep the exposition concise.

The main contribution of Mihm and Ozbek (2019) is to show when $\succsim$ can be rationalised uniquely by $(\succsim_H)$, that is, when future preferences can be anticipated unambiguously. To do so, they assume X_t is a space of lotteries endowed with the Euclidean metric and a first order stochastic dominance relation $\rhd$.[2] Such $\rhd$ must satisfy some regularity conditions, for which we have to introduce a few more definitions. First, $\succsim$ is $\rhd$-sensitive if $x \rhd y \implies (h,x,a) \nsim (h,y,a)$ $\forall h \in H$, $x,y \in X_{|h|}$ and $a \in \mathcal{X}_{|h|+1}$. $\rhd$-sensitivity means that given a history h and a continuation problem a, the DM is not indifferent between two lotteries if one first order stochastically dominates the other. This is essentially saying that first order stochastic dominance is relevant for the ranking of alternatives. Second, $(\succsim_H)$ is $\rhd$-regular if all $\succsim_h \in (\succsim_H)$ satisfy:

(i) Monotonicity: $\forall x,y \in X_{|h|}$ with $x \neq y$, $\forall a \in \mathcal{X}_{|h|+1}$, $x \rhd y \implies (x,a) \succ_h (y,a)$;

(ii) Upper semicontinuity: $\forall x \in X_{|h|} \times \mathcal{X}_{|h|+1}$, $\forall a,b \in \mathcal{X}_{|h|+1}$, $\{y \in X_{|h|} : (x,a) \succ_h (y,b)\}$ is open.

Monotonicity means that $\succsim_h$ ranks alternatives consistently with first order stochastic dominance, while upper semicontinuity ensures that preferences are well-behaved, i.e., they can be maximised. Finally, combining these definitions, we define a consistent preference $\succsim$ to be regular if it is $\rhd$-sensitive and it is compatible with a plan consistent with a $\rhd$-regular ranking system $(\succsim_H)$. Put simply, $\succsim$ is regular when first order stochastic dominance is relevant for the ranking of alternatives and when it is well-behaved.

In order to characterise a consistent and $\rhd$-sensitive $\succsim$ as being rationalised by a $\rhd$-regular ranking system $(\succsim_H)$, Mihm and Ozbek (2019) impose a regularity assumption on $\succsim$ that mirrors $\rhd$-regularity. Although the characterisation of regular $\succsim$ is an important result *per se*, the most important theorems relate to the elicitation of the ranking system $(\succsim_H)$ that rationalises $\succsim$. Consider the revealed ranking system $(\succsim^*_H)$ induced by $\succsim$ as follows: $\forall h \in H$, $x,y \in X_{|h|}$ and $a,b \in \mathcal{X}_{|h|+1}$, if $\exists z \in X_{|h|}$ such that $z \rhd y$ and $(h, \{(x,a),(z,b)\}) \nsim (h,z,b)$ then $(x,a) \succ^*_h (y,b)$, while otherwise $(y,b) \succsim^*_h (x,a)$. This is telling us that when the DM finds, from a time 0 perspective, an alternative (x,a) more attractive than (z,b) at history h, and $z \rhd y$, they anticipate preferring (x,a) to (y,b) when history h realises, that is, $(x,a) \succ^*_h (y,b)$.

Crucially, Mihm and Ozbek (2019) show that if a $\rhd$-regular ranking system $(\succsim_H)$ rationalises a consistent and $\rhd$-sensitive $\succsim$, then $(\succsim_H) = (\succsim^*_H)$. This theorem is fundamental to the extent that it states that it is possible to anticipate future preferences starting from today's preferences and a dominance relation. Moreover, Mihm and Ozbek (2019) show that if we use two different dominance relations which induce two ranking systems rationalising $\succsim$, then the two ranking systems must coincide. Hence, inference about future preferences is not dependent on the choice of the dominance relation used to elicit them. In summary, Mihm and Ozbek (2019) provide a set of regularity conditions under which it is possible to anticipate unambiguously future rankings. Armed with the knowledge of a framework in which it is possible to make sharp predictions about future preferences, we now move on to analyse papers that allow us to test whether and to what extent agents are naive.

Sophistication and preferences for commitment

Noor (2011) presents an infinite horizon model in which choices are driven by a struggle between normative and temptation preferences, and sophisticated agents have a preference for commitment as it allows them to align actual with desirable behaviour. Consider a compact metric space of consumption alternatives C and a space of infinite horizon menus Z. The primitive in Noor (2011) is a closed- and non-empty-valued choice correspondence $\mathcal{C} : Z \rightrightarrows \Delta(C \times Z)$, interpreted as the DM facing a menu $z \in Z$ in period t and choosing a lottery $\mu \in \mathcal{C}(z)$, that is,

[2] Actually, Mihm and Ozbek (2019) have a more general definition of $\rhd$ and its properties, but first order stochastic dominance defined on a space of lotteries satisfies them.

choosing a distribution over consumption and next period menus. The choice correspondence has a U-V representation if for all $z \in Z$ it satisfies

$$\mathcal{C}(z) = \arg\max_{\mu \in z}\{U(\mu) + V(\mu)\}$$

where $U, V : \Delta(C \times Z) \to \mathbb{R}$ are such that $U(\mu) = \int_{C\times Z} u(c) + \delta W(z)\, \mathrm{d}\mu$ and $V(\mu) = \int_{C\times Z} v(c) + [\beta W(z) + \gamma \max_{\eta \in z} V(\eta)]\, \mathrm{d}\mu$, with $\mu \in \Delta(C \times Z)$. δ, β and γ are parameters that determine the kind of temptation experienced by the DM.

The normative expected utility U aggregates a Bernoulli instantaneous utility u and a discounted utility of continuation menus W. $W : Z \to \mathbb{R}$ is introduced in Gul and Pesendorfer (2004) and is a value function corresponding to utility net of self-control costs due to temptation. The temptation expected utility V, similarly, aggregates a Bernoulli instantaneous utility v, a discounted utility of continuation menus W, and a discounted pure temptation value $\max_{\eta \in z} V(\eta)$ of menu z. A menu z with high pure temptation value will induce the DM to assign more weight to it when choosing the optimal μ. The U-V representation proposed by Noor (2011) formalises the struggle between normative and temptation utilities, by stating that the DM chooses from menu z so to maximise the sum of the two conflicting utilities.

The main contribution of Noor (2011) is to provide an axiomatic foundation of the U-V representation. Among the various axioms, we now discuss the one relating to sophistication. The choice correspondence defines a preference relation $\succsim$ on $\Delta(C \times Z)$ as follows: $\mu \succsim \eta \iff \mu \in \mathcal{C}(\{\mu, \eta\})$, that is, μ is preferred to η if μ would be chosen from the binary menu $\{\mu, \eta\}$. Noor (2011) denotes by $\{\mu\}^{+t}$ the choice that yields a fixed level of consumption $\bar{c} \in C$ for all periods between 0 and t and prescribes choosing lottery μ from the menu faced in period t (with $\{\mu\}^{+0} = \mu$). This formulation allows to compare alternatives in a menu faced t periods ahead, net of the value of the consumption stream leading to that menu. Moreover, Noor (2011) introduces the fundamental concept of distancing: $\{\mu\}^{+t} \succsim \{\eta\}^{+t}$ is interpreted as μ being preferred to η when the two alternatives are evaluated from a distance of t periods. The normative expected utility U represents the preference that compares alternatives from an infinite distance, that is, when the DM ranking alternatives is fully distanced from the consequences of their choices.

Noor (2011)'s sophistication axiom states that for all $\mu, \eta \in \Delta(C \times Z)$ and for all $t \geq 0$, if $\{\mu\}^{+t} \succ \{\eta\}^{+t}$, then $\{\mu, \eta\}^{+t} \succ \{\eta\}^{+t} \iff \mu \succ \eta$. Notice that $\{\mu\}^{+t} \succ \{\eta\}^{+t}$ means that the DM thinks that t periods in the future they will prefer μ to η. Sophistication requires that anticipating to choose μ after t periods ($\{\mu, \eta\}^{+t} \succ \{\eta\}^{+t}$) is equivalent to actually choosing μ ($\mu \succ \eta$). In other words, sophistication amounts to the ability of correctly anticipating future choices. Notice that sophistication does not mean that preferences cannot be reverted. Indeed, dynamic inconsistency is contemplated by the model as it is possible that $\{\mu\}^{+t} \succ \{\eta\}^{+t}$ and $\{\eta\}^{+t'} \succsim \{\mu\}^{+t'}$ ($t \neq t'$). Sophistication, however, requires judgements made from a distance of t periods to take into account the temptation that might arise between the *ex ante* judgement and the *ex post* choice. In this context, sophisticated agents have a preference for commitment, as they correctly anticipate the evolution of preferences. One of the merits of Noor (2011) is to be among the first to provide a testable definition of sophistication, and to clarify the link between sophistication and time consistent behaviour.

Measuring absolute and comparative naivete

Ahn et al. (2019) improve on testability and measurement of sophistication by providing behavioural definitions of absolute and comparative naivete. Let $\Delta(C)$ denote the set of lotteries over consumption alternatives, and $\mathcal{K}(\Delta(C))$ a collection of menus, that is, a family of non-empty compact subsets of $\Delta(C)$. The primitives of the model are a preference relation

$\succsim$ on $\mathcal{K}(\Delta(C))$, describing the ranking of menus before the experience of temptation, and a choice function $\mathcal{C} : \mathcal{K}(\Delta(C)) \to \Delta(C)$, encoding choice behaviour after the experience of temptation. Given an expected utility function $u : \Delta(C) \to \mathbb{R}$ and a menu $z \in \mathcal{K}(\Delta(C))$, we write $B_u(z) = \arg\max_{p\in z} u(p)$.

Ahn et al. (2019) define a Strotz representation of the primitives $(\succsim, \mathcal{C})$ as a triple $(u, v, \hat{v})$ of non-constant expected utilities such that $U : \mathcal{K}(\Delta(C)) \to \mathbb{R}$ defined by $z \mapsto \max_{p\in B_{\hat{v}}(z)} u(p)$ represents $\succsim$ and $\mathcal{C}(z) \in B_u(B_v(z))$ for all $z \in \mathcal{K}(\Delta(C))$. A Strotzian DM *ex ante* maximises $\hat{v}$, but after the experience of temptation their *ex post* choice maximises v. This is the representation on which Ahn et al. (2019) characterise absolute and comparative naivete.

Ahn et al. (2019) define a DM to be sophisticated if for all menus $z \in \mathcal{K}(\Delta(C))$, $z \sim \{\mathcal{C}(z)\}$. This means that a sophisticated individual is *ex ante* indifferent between being presented the menu z or the singleton menu $\{\mathcal{C}(z)\}$ whose only alternative is the one chosen *ex post*. Conversely, a DM is naive if $z \succsim \{\mathcal{C}(z)\}$ for all menus $z \in \mathcal{K}(\Delta(C))$. A naive DM choosing $\mathcal{C}(z)$ *ex post* thinks *ex ante* they will choose more virtuously than $\mathcal{C}(z)$ from menu z and, thus, prefer retaining the flexibility of the whole menu rather than committing to the *ex post* choice.

Absolute naivete is characterised by Ahn et al. (2019) through the relative alignment with u of $\hat{v}$, the *ex ante* anticipated temptation preference, and v, the *ex post* realised temptation preference. Writing $u \approx v$ when u is a positive affine transformation of v, Ahn et al. (2019) define $\hat{v}$ to be more u-aligned than v, written $\hat{v} \gg_u v$, if $\hat{v} \approx \alpha u + (1-\alpha)v$ for some $\alpha \in [0, 1]$. Ahn et al. (2019)'s characterisation shows that the DM is naive if and only if $\hat{v} \gg_u v$ and sophisticated if and only if $\hat{v} \approx v$. The interpretation is simple: a naive DM *ex ante* anticipates maximising $\hat{v}$ under temptation, but ends up maximising v instead, the latter being further away from (less aligned with) the normative utility u.

Next, Ahn et al. (2019) introduce a notion of comparative naivete, characterised by under-demand for commitment. Specifically, they define DM 1 to be more naive than DM 2 if for all menus $z \in \mathcal{K}(\Delta(C))$ and for all lotteries $p \in \Delta(C)$, $z \succ_2 \{p\} \succ_2 \{\mathcal{C}(z)\} \implies z \succ_1 \{p\} \succ_1 \{\mathcal{C}(z)\}$. DM 2 is naive as they would be *ex post* better off by committing to $\{p\}$, but they refuse such commitment as they naively believe $z \succ_2 \{p\}$. DM 1 is more naive as whenever DM 2 displays such naive behaviour, DM 1 behaves naively as well. Comparative naivete is again characterised through the concept of alignment. Ahn et al. (2019) show that if DM 1 and DM 2 have Strotz representations $(u, v_1, \hat{v}_1)$ and $(u, v_2, \hat{v}_2)$, then DM 1 is more naive than DM 2 if and only if either $\hat{v}_2 \approx v_2$ (DM 2 is sophisticated) or $\hat{v}_1 \gg_u \hat{v}_2 \gg_u v_2 \gg_u v_1$.

Notice that here we are not only requiring DM 1 to be more optimistic than DM 2 about future behaviour ($\hat{v}_1 \gg_u \hat{v}_2$), but also that DM 1 makes worse *ex post* choices than DM 2 ($v_2 \gg_u v_1$). A more intuitive version of this characterisation would simply state that $\hat{v}_1$ and v_1 are more distant than $\hat{v}_2$ and v_2, where distance should be understood in terms of u-alignment. This is related to the concept of overvaluation of menus also introduced by Ahn et al. (2019) to characterise comparative naivete. However, we do not discuss this in the present work as it seems less compelling (given that it has a cardinal interpretation) and since overvaluation is not equivalent to (but is implied by) under-demand for commitment in the more general setting with a random Strotz representation.

Motivated naivete

If Ahn et al. (2019) presented definitions of absolute and relative naivete, Tang and Zhang (2023) focus on modelling how naivete of agents can be motivated, thus giving a different persepctive on tests for comparative naivete. Let $\Delta(X)$ be a space of lotteries and $\mathcal{M}$ denote the collection of all subsets of $\Delta(X)$, that is, all possible menus. $\mathcal{U}$ is a compact set of utility indexes $u : X \to \mathbb{R}$ with which we can represent (in an expected utility fashion) all possible preferences over lotteries. Then, for any non-empty and compact collection of utility indexes $\mathcal{V} \subseteq \mathcal{U}$ and for any menu $A \in \mathcal{M}$, Tang and Zhang (2023) define $c(\mathcal{V}, A) := \{p \in A : \exists v \in$

$\mathcal{V}$ s.t $v(p) = \max_{q \in A} v(q)\}$, which is the set of lotteries in the menu A whose choice can be rationalised by some utility index in $\mathcal{V}$.

A preference $\succsim$ over menus is said to be a motivated naive menu preference if there is a $u \in \mathcal{U}$ and a non-empty and compact $\mathcal{V} \subseteq \mathcal{U}$ such that for all $A, B \in \mathcal{M}$, $A \succsim B \iff \max_{p \in c(\mathcal{V}, A)} u(p) \geq \max_{p \in c(\mathcal{V}, B)} u(p)$. The interpretation is that the DM holds a current (or normative) preference u and believes that in the future they will hold preferences within the set $\mathcal{V}$. The DM evaluates a menu A by considering the lotteries whose choice from the menu can be rationalised by future preferences, and picking from these the lottery that maximises current preferences. The menu preference is motivated since the set $\mathcal{V}$ is constructed based on past behaviour; however the menu preference is naive to the extent that the value of the menu is determined by current preferences which might not be aligned with the preferences that will realise in the future. In this case we say that $\succsim$ is represented by $(u, \mathcal{V})$.

The main contribution of Tang and Zhang (2023) is providing an axiomatic foundation of motivated naive menu preferences. The axioms characterise a preference $\succsim$ that is represented by $(u, \mathcal{V})$, where u is unique. On the other hand, the set of future preferences is not necessarily unique, i.e., the same preference can be represented by $(u, \mathcal{V})$ and $(u, \widehat{\mathcal{V}})$. In order to tackle this issue, Tang and Zhang (2023) invoke the concept of u-alignment used in Ahn et al. (2019) as well. When v is more u-aligned than v' ($v \gg_u v'$), we have that $u(p) \geq u(q)$, with $p \in c(\{v\}, A)$ and $q \in c(\{v'\}, A)$ $\forall A \in \mathcal{M}$. Hence, the value of a menu A ($\max_{p \in c(\mathcal{V}, A)} u(p)$) is determined solely by the $v \in \mathcal{V}$ that are $\gg_u$-undominated in $\mathcal{V}$, that is, the preferences that are mostly aligned with u. The second contribution of Tang and Zhang (2023) is to show that when $\succsim$ is represented by $(u, \mathcal{V})$ and $(u, \widehat{\mathcal{V}})$, the sets of $\gg_u$-undominated preferences in $\mathcal{V}$ and $\widehat{\mathcal{V}}$ are identical. This means that, despite the set of future preferences is not necessarily unique, the set of future preferences that are relevant for the assessment of a menu is unique.

Finally, Tang and Zhang (2023) provide a behavioural characterisation of comparative naivete. $\succsim_1$ is defined to be more naive than $\succsim_2$ if $\forall A \in \mathcal{M}$ and $p \in \Delta(X)$, we have $A \succsim_2 p \implies A \succsim_1 p$. Similar to other definitions of relative naivete discussed earlier, here the idea is that whenever $\succsim_2$ prefers the flexibility of menu A to the commitment to lottery p, also a more naive $\succsim_1$ will prefer flexibility over commitment. In order to characterise comparative naivete, Tang and Zhang (2023) extend the notion of u-alignment to sets of functions: $\mathcal{V}_1 \gg_u \mathcal{V}_2$ if $\forall v_2 \in \mathcal{V}_2$, $\exists v_1 \in \mathcal{V}_1$ such that $v_1 \gg_u v_2$. Then, given preferences $\succsim_1$ and $\succsim_2$ represented by $(u_1, \mathcal{V}_1)$ and $(u_2, \mathcal{V}_2)$ respectively, Tang and Zhang (2023) show that $\succsim_1$ is more naive than $\succsim_2$ if and only if $u_1 = u_2 =: u$ and $\mathcal{V}_1 \gg_u \mathcal{V}_2$. In words, comparative naivete is described by the alignment of the set of possible future preferences with the normative preference.

Conclusion

In conclusion, we have discussed few papers conceptualising naivete of agents. Mihm and Ozbek (2019) provided a framework in which current preferences and a dominance relation are sufficient to make inference about future preferences. Comparing *ex ante* and *ex post* choices would then allow us to understand whether and to what extent agents are sophisticated. Such tests can be carried out by directly inspecting whether the behaviour of agents is consistent with the behavioural definitions of absolute and comparative naivete given by Noor (2011), Ahn et al. (2019) and Tang and Zhang (2023). Then, through the representations of absolute and comparative naivete provided by Ahn et al. (2019) and Tang and Zhang (2023), it is possible to describe naivete through specific utility representations, thus making the concept more tractable. Overall, these are important tools through which it becomes possible to model inter-temporal decision making. Future research could focus on understanding whether agents become less naive when facing a decision problem repeatedly, and whether naivete is context dependent, that is, whether agents behave naively in some contexts and are instead sophisticated in others.

References

Ahn, David S., Ryota Iijima, Yves Le Yaouanq, and Todd Sarver, "Behavioral Characterizations of Naivete for Time-Inconsistent Preferences," *The Review of Economic Studies*, 2019, *86*, 2319–2355.

Gul, Faruk and Wolfgang Pesendorfer, "Self-Control and the Theory of Consumption," *Econometrica*, 2004, *72* (1), 119–158.

_ **and** _ , "The Revealed Preference Theory of Changing Tastes," *The Review of Economic Studies*, 2005, *72* (2), 429–448.

Mihm, Maximilian and Kemal Ozbek, "On the Identification of Changing Tastes," *Games and Economic Behavior*, 2019, *116*, 203–216.

Noor, Jawwad, "Temptation and Revealed Preference," *Econometrica*, 2011, *79*, 601–644.

Tang, Rui and Mu Zhang, "Motivated Naivete," *Journal of Economic Theory*, 2023, *209*, 105636.