Liberal parentalism

A. Heifetz, E. Minelli & H. Polemarchakis

(This paper also appears as CRETA Discussion Paper No: 59)

This paper has been updated & accepted for publication in Journal of Public Economic Theory

July 2020 No: 1289

Warwick Economics Research Papers

ISSN 2059-4283 (online)

ISSN 0083-7350 (print)

Liberal parentalism

A. Heifetz[1] E. Minelli[2] H. Polemarchakis[3]

July 1, 2020

[1]Open University of Israel, aviadhe@openu.ac.il
[2]University of Brescia, enrico.minelli@unibs.it
[3]University of Warwick, H.Polemarchakis@warwick.ac.uk

Abstract

What normative constraints should bind parents (or policy makers) if they intervene in the choices of children (or constituencies) whose preferences evolve over time? For a sophisticated child who anticipates correctly his preference change, we prove that *generically* there exist parental interventions that are Pareto improving over the backward induction path that the child will follow on his own. If, in contrast, the child misperceives his future preferences, Pareto improving interventions might not exist, and even *nudges* might be painfully sobering. The parent may then choose to *minimize the maximal disappointment along time* that her benevolent intervention would cause.

Parents, in the wide sense of the term, should want their children to be happy. This is so within nuclear families, as well as between governments and their constituencies. To this effect, parents may convey to their children information that the children do not have in the first place. But after all is said, should parents, if they can, intervene in their children's choices to make them happier?

If children's preferences over action paths do not change over time, there is no such need: by the principle of optimality in dynamic programming, a child will follow from one period to the next the same overall plan that already from the very start he would like there to be implemented, and in particular no parental intervention can further enhance the child's happiness.

But what if the child's preferences over action paths *do* change over time? In each period, based on his current belief about his future preferences and beliefs, the child would anticipate his own future reactions[1] to any choice he can make today, and then make a choice that together with these anticipated reactions would be optimal according to his current preferences. With a finite horizon, like when choosing education or vocational training before adulthood, or saving up to retirement, this is done, implicitly if not explicitly, by backward induction. In other words, the choices along time would constitute a *Strotz-Pollak equilibrium* (Strotz 1956, Pollak 1968, Goldman 1980), i.e. a subgame-perfect equilibrium (SPE) in the 'intrapersonal game' (Laibson 1997) among the selves of the child across the time periods.

The mature and experienced parent, in contrast, may forecast the preference evolution more accurately than the child. In addition, she may have her own perspective on how to 'responsibly' average or aggregate the child's evolving preferences along time into one preference relation over action plans, so as to balance youthful vivacity with adult thriving. The optimal plan according to this aggregated preference relation may very well differ from the path that the child will follow by backward induction on his own.

So the normative question poses itself once again[2]: if the parent can, should she intervene and induce the optimal plan according to her overall, aggregate view of the child's evolving preferences? The question is accentuated by the fact that if the parent *can* intervene, non-intervention becomes one out of many possible decisions that the parent can make, and as such non-intervention need not necessarily be considered to be normatively neutral.

Some of the literature thus far (Phelps and Pollak 1968, Laibson 1997, Jackson and Yariv 2014, 2015, Kang 2019, Kang and Ye 2019) suggests that the parent should restrict herself to interventions that are *Pareto-improving*, i.e. to inducing action paths that the child deems at least as good as the backward induction

[1] Or, more generally, when the future is uncertain, the *distribution* of his own future reactions.

[2] See e.g. Ericson and Laibson 2019, open question 9.

path in each and every period, and strictly better in at least one period. Such an intervention may be thought of as normatively non-controversial, because in all periods the child will welcome such a change, enforced by the commitment power of the parent that the child himself lacks. The parent may then decide to induce the plan which is optimal according to her aggregated preferences *subject* to the Pareto-improvement constraint.

With this perspective, it is important to know whether Pareto-improving interventions exist in the first place. In section 2 we therefore present a framework that makes explicit that even though bygones are bygones, current well-being may well depend also on remembered, past experience. In this framework we present a general result: when the child is *sophisticated,* i.e. when in each period he anticipates correctly his own future preference change, then Pareto-improving interventions *generically* exist. That is, if for some preference evolution along time, the subgame-perfect equilibrium path that a sophisticated child will follow on his own so happens to be Pareto optimal, then with slightly perturbed preferences his subgame-perfect equilibrium path will no longer be Pareto optimal, and, moreover, neither will it be with further, smaller perturbations of the child's preferences. In other words, a sophisticated child chooses Pareto optimally by himself only with knife-edge profiles of preferences[3].

A more realistic assumption, though, is that a child with evolving preferences is not perfectly sophisticated, but rather at least partially naive (O'Donoghue and Rabin 1999), that is uncertain about his own future preferences, if not outright wrong about them. In section 3 we therefore define a simple but general type-space framework to encompass also such forms of naivité. *Nudges* (Thaler and Sunstein 2008), or more generally interventions that are purely informational, are modeled in this type-space framework by a change in the state of the world that expresses the new beliefs the child would then hold across the time periods.

We then show by example that with at least some naivité, even when the optimal path from the parent's perspective differs from the backward induction path, a Pareto-improving intervention might not exist, neither by inducing a different path nor by a nudge. What should the parent do in such cases?

In section 4 we propose one possible answer, namely a normative approach by which the parent should only intervene in a way that minimizes the maximal disappointment of the child across the time periods, relative to the backward induction benchmark. We show by way of example that subject to this minimax constraint, there may indeed exist such an intervention that enhances the child's aggregated well-being from the parent's perspective.

In fact, this normative approach may be applied also to cases in which Pareto-

[3]In particular, stable preferences which do not change over time is one such knife-edge case, in which by the principle of optimality a sophisticated child chooses Pareto optimally.

improving interventions *do* exist. In these cases, the normative criterion would call the parent to choose only among interventions that maximize the minimal well-being enhancement across the time periods. We show, again by way of example, how this may (not surprisingly) alter the optimal path that the parent would induce given this additional constraint, in comparison with her intervention subject only to the Pareto-improvement constraint.

We conclude in section 6 with a discussion. Example details appear in the appendix, and the genericity proof is in the online appendix.

1 Sophisticated children rarely choose optimally on their own

For each period $i = 1, ..., n$, let $X_i \subset \mathbb{R}^{k_i}$ be the compact, convex choice set of the child at period i, henceforth dubbed 'self i'. A current self cannot 'pre-program' future selves' choices.[4]

Denote $X = X_1 \times ... \times X_n$, and $X_{\leq i} = \prod_{j \leq i} X_j$ the space up-to-i initial paths.[5] Let $u_i : X \to \mathbb{R}$ be the utility function of self i, where $u_i \in C^2(X, \mathbb{R})$, the space of twice continuously differentiable functions, endowed with the topology of the norm $\|u\| = \sup_{x \in X} \{\|u(x)\|, \|Du(x)\|, \|D^2 u(x)\|\}$. The space of utility profiles,

$$u = (u_1, ..., u_n) \in \mathcal{U} \equiv \left(C^2(X, \mathbb{R})\right)^n$$

is endowed with the product topology. *Sophistication* means that with a utility profile u, implicitly each self i anticipates correctly the future selves' utility functions $u_{i+1}, ..., u_n$, knows that each future self $j > i$ will also anticipate correctly its own future selves' utility functions $u_{j+1}, ..., u_n$, etcetera.[6]

Without intervention, a sophisticated child is assumed to choose by backward induction, i.e. to follow the path $\hat{x} = (\hat{x}_1, ..., \hat{x}_n)$ induced by a Strotz-Pollak equilibrium of best replies $(b_i : X_{\leq i-1} \to X_i)_{i=1}^n$, which is a subgame-perfect equilibrium of the perfect-information dynamic game between the selves.[7] At such an

[4] Put differently, whatever device a particular self *does* have, if at all, for influencing future selves' behavior, choosing to influence future behavior in such a way is encoded within the available choices $x_i \in X_i$, and by the utility functions (as defined next) of future selves $j > i$, that depend in particular on x_i.

[5] Or, more generally, $X \subseteq X_1 \times ... \times X_n$ and $X_{\leq i} \subseteq \prod_{j \leq i} X_j$ – in case the past choice path $(x_1, ..., x_{j-1})$ of the selves $k < j$ may limit the feasible choices of self j.

[6] In section 3 below we will present an extended framework where this assumption can be made explicit, a framework that also allows for various deviations from such perfect sophistication.

[7] For proofs of equilibrium existence see Harris (1985), Hellwig and Leininger (1987), Hellwig et al. (1990), Alós-Ferrer and Ritzberger (2016). If best replies are not unique, breaking ties in different ways may lead to multiple equilibria.

equilibrium,

$$\begin{aligned}
&b_n\left(x_{\leq n-1}\right) \in \arg\max\nolimits_{X_n} u_n\left(x_{\leq n-1}, \cdot\right), \\
&b_{n-1}\left(x_{\leq n-2}\right) \in \arg\max\nolimits_{X_{n-1}} u_{n-1}\left(x_{\leq n-2}, \cdot, b_n\left(x_{\leq n-2}, \cdot\right)\right), \\
&b_{n-2}\left(x_{\leq n-3}\right) \in \arg\max\nolimits_{X_{n-2}} u_{n-2}\left(x_{\leq n-3}, \cdot, b_{n-1}\left(x_{\leq n-3}, \cdot\right),\right. \\
&\left. b_n\left(x_{\leq n-3}, \cdot, b_{n-1}\left(x_{\leq n-3}, \cdot\right)\right)\right), \\
&\vdots \\
&b_1\left(\emptyset\right) \in \arg\max\nolimits_{X_1} u_1\left(\cdot, b_2\left(\cdot\right), b_3\left(\cdot, b_2\left(\cdot\right)\right), \ldots\right).
\end{aligned}$$

The corresponding backward induction path $\hat{x} = (\hat{x}_1, ..., \hat{x}_n)$ is then defined inductively by

$$\begin{aligned}
&\hat{x}_1 = b_1\left(\emptyset\right), \\
&\hat{x}_2 = b_2\left(\hat{x}_1\right), \\
&\hat{x}_3 = b_3\left(\hat{x}_1, \hat{x}_2\right), \\
&\vdots \\
&\hat{x}_n = b_n\left(\hat{x}_1, \hat{x}_2, ..., \hat{x}_{n-1}\right).
\end{aligned}$$

It yields the utility levels $\hat{u} = (\hat{u}_1, ..., \hat{u}_n)$ where $\hat{u}_i = u_i(\hat{x})$.

The parent has an average/aggregation function $V : \mathbb{R}^n \to \mathbb{R}$ over the utility levels of the different selves $i = 1, ..., n$. This defines, indirectly, the utility function of the parent $v : X \to \mathbb{R}$ over choice paths, $v(x) = V(u_1(x), ..., u_n(x))$.

For simplicity we assume that the parent can oblige the child to follow any path $x \in X$ that the parent likes.[8] Still, the parent may like to limit herself to a subset of paths with some normatively desirable properties.

A path $x \in X$ is a called a Pareto improvement over a path $\bar{x} \in X$ if $u_i(x) \geq u_i(\bar{x})$ for every $i = 1, ..., n$, and $u_j(x) > u_j(\bar{x})$ for some $1 \leq j \leq n$. Suppose, first, that the parent would like to limit herself to inducing the child only to choice paths x that are Pareto improvements over the backward induction path $\hat{x}$ that the child would follow on his own.

Example 1. A sophisticated child has to finish a chore of size 3 in three days $i = 1, 2, 3$. Carrying out quantity x_i of the chore on day i takes x_i^2 hours, and the child enjoys his remaining leisure time $24 - x_i^2$.[9] The child's time preference is captured by $\beta - \delta$ quasi-hyperbolic discounting, with $\beta = \frac{1}{2}$ and $\delta = 1$. At days $2, 3$ the child's memory of his past leisure augments his well-being, with no

[8]We note that leading the child to follow a particular path x need not necessarily involve direct coercion by the parent. Instead, the parent may be able to provide *incentives* for the child to follow x, e.g. via taxes or subsidies accompanied by an unconditional lump-sum transfer, as in Kang (2019).

[9]This example is similar to the 'cake-eating' problem under changing tastes, as analyzed e.g. in Goldman (1979). Here, in contrast, the felicity from the good (leisure) is linear whereas the 'cost' of leisure (time put aside for the chore) is convex.

discounting. Thus, if the child's division of the chore along days $i = 1, 2, 3$ is $x = (x_1, x_2, x_3)$, his corresponding selves' utility functions are

$$\begin{aligned} u_1(x_1, x_2, x_3) &= (24 - x_1^2) + \tfrac{1}{2}(24 - x_2^2) + \tfrac{1}{2}(24 - x_3^2), \\ u_2(x_1, x_2, x_3) &= (24 - x_1^2) + (24 - x_2^2) + \tfrac{1}{2}(24 - x_3^2), \\ u_3(x_1, x_2, x_3) &= (24 - x_1^2) + (24 - x_2^2) + (24 - x_3^2). \end{aligned}$$

Across the three days the parent values equally the child's happiness[10], leading to the parent's utility function

$$v(x_1, x_2, x_3) = \frac{1}{3}\sum_{i=1}^{3} u_i(x_1, x_2, x_3) = \left(24 - x_1^2\right) + \frac{5}{6}\left(24 - x_2^2\right) + \frac{2}{3}\left(24 - x_3^2\right).$$

With the constraint $x_3 = 3 - x_1 - x_2$, one can verify (see the details in the appendix) that

$$\bar{x} = \arg\max_{x_1, x_2} v(x_1, x_2, 3 - x_1 - x_2)$$

is not a Pareto improvement over the backward induction path $\hat{x}$.

However, Pareto improvements over $\hat{x}$ do exist:

(a) The child would be happy to work somewhat more on day 1 if only he could know that he will be bound to work somewhat more also on day 2, and not leave so much of the chore to the last day, as he correctly anticipates on day 1 that he would be doing on day 2 without intervention;

(b) On day 2 the child would be happy to be committed to a package deal in which he works somewhat more on both days 1 and 2; and

(c) On day 3 the child would definitely be happy if less of the largest share of the chore is left for him to finish.

Moreover, the Pareto improving path $\breve{x}$ that maximizes v satisfies $v(\breve{x}) > v(\hat{x})$, so the parent would like to intervene and induce $\breve{x}$ (even though implementing the non-Pareto improving $\bar{x}$ would entail an even higher utility, $v(\bar{x}) > v(\breve{x})$).

A path $\tilde{x} \in X$ is called *Pareto optimal* if there does not exist a Pareto improvement $x \in X$ over $\tilde{x}$.

Theorem. There is an open and dense subset of utility profiles $\mathcal{U}_0 \subseteq \mathcal{U}$ of sophisticated children, for whom no interior backward induction path $\hat{x}$ is Pareto optimal.

The proof, based on a transversality argument, appears in the online appendix.

[10] *Not* the child's leisure time – that would have amounted to a complete identification of the parent with the child's perspective on day 3, waving off the child's perspectives on days 1 and 2.

Thus, for sophisticated children with *typical* or *generic* utility profiles, namely utility profiles in $\mathcal{U}_0$, there exists a Pareto improving path $x \in X$ over any interior backward induction path $\hat{x}$. If the parent can induce the child to follow a Pareto improving x instead of $\hat{x}$, this will be weakly preferred by the child in all time periods, and strictly preferred in some time period(s). Moreover, if the parent's utility is some weighted average of the child's utility levels at the different time periods, then also the parent will prefer to induce x instead of letting the child follow the backward induction path $\hat{x}$ on his own.

2 A type space for possible misperceptions of future preferences

But children are not always sophisticated, and might anticipate their future preferences, and therefore their future choices, differently than they will actually be. Examples of such naivité or partial naivité were discussed e.g. by O'Donoghue and Rabin (1999). The following type space framework is intended to capture such states of affairs in general.

For each self $i = 1, ..., n$, let T_i be a measurable space of self i's types. Types $\tau_i \in T_i$ are measurably associated with

1. a belief (i.e. a probability measure) β_{τ_i} on the states of the world $T \subseteq \prod_{j=1}^{n} T_j$, where β_{τ_i} features perfect recall, i.e. has the property that in each state $\tau = (\tau_1, ..., \tau_n) \in T$, $\beta_{\tau_i}\left(\{(\tau_1, ..., \tau_i)\} \times \prod_{j=i+1}^{n} T_j\right) = 1$,[11,12]

2. a measurable utility function $u_{\tau_i} : X \to \mathbb{R}$, that not only represents preferences over choice paths $x \in X$, but also meaningfully expresses well-being

[11] So when no confusion may arise, we will interchangeably refer to β_{τ_i} as a belief on $\prod_{j=i+1}^{n} T_j$.

[12] The space of beliefs μ on a measurable space Y is endowed with the σ-algebra generated by the sets of the form $\{\mu : \mu(E) \geq p\}$ for $p \in [0, 1]$ and measurable events $E \subseteq Y$.

comparably across selves,[13,14] and

3. a measurable best-reply function $b_{\tau_i} : X_{\leq i-1} \to X_i$, where

$$b_{\tau_n}\left(x_{\leq n-1}\right) \in \arg\max_{X_n} u_{\tau_n}\left(x_{\leq n-1}, \cdot\right),$$
$$b_{\tau_{n-1}}\left(x_{\leq n-2}\right) \in \arg\max_{X_{n-1}} \int_{T_n} u_{\tau_{n-1}}\left(x_{\leq n-2}, \cdot, b_{\tilde{\tau}_n}\left(x_{\leq n-2}, \cdot\right)\right) \mathrm{d}\beta_{\tau_{n-1}}\left(\tilde{\tau}_n\right),$$
$$b_{\tau_{n-2}}\left(x_{\leq n-3}\right) \in \arg\max_{X_{n-2}} \int_{T_{n-1}\times T_n} u_{\tau_{n-2}}\Big(x_{\leq n-3}, \cdot, b_{\tilde{\tau}_{n-1}}\left(x_{\leq n-3}, \cdot\right)$$
$$b_{\tilde{\tau}_n}\Big(x_{\leq n-3}, \cdot, b_{\tilde{\tau}_{n-1}}\left(x_{\leq n-3}, \cdot\right)\Big)\Big) \mathrm{d}\beta_{\tau_{n-2}}\left(\tilde{\tau}_{n-1}, \tilde{\tau}_n\right),$$
$$\vdots$$
$$b_{\tau_1}\left(\emptyset\right) \in \arg\max_{X_1} \int_{T_2\times\ldots\times T_n} u_{\tau_1}\left(\cdot, b_{\tilde{\tau}_2}\left(\cdot\right), b_{\tilde{\tau}_3}\left(\cdot, b_{\tilde{\tau}_2}\left(\cdot\right)\right), \ldots\right) \mathrm{d}\beta_{\tau_1}\left(\tilde{\tau}_2, \ldots, \tilde{\tau}_n\right),$$

for $i = n, ..., 1$.[15,16]

For (a particularly simple) example, each utility profile $u = (u_1, ...u_n)$ of a sophisticated child, together with a Strotz-Pollak equilibrium $(b_i : X_{\leq i-1} \to X_i)_{i=1}^{n}$

---

[13]To the latter effect, in the background there may e.g. be for each type τ_j of self j an instantaneous felicity function $f_{\tau_j} : X_j \to \mathbb{R}$, so that

$$u_{\tau_i}\left(x_1, ..., x_n\right) = \sum_{j=1}^{n} \omega_{\tau_i}\left(j\right) \left(\int_T \left(f_{\tau_j}\left(x_j\right)\right) \mathrm{d}\beta_{\tau_i}\right),$$

for some positive weights $\omega_{\tau_i}(j)$ (these weights may represent, for example, time discounting by the type τ_i) or, more generally,

$$u_{\tau_i}\left(x_1, ..., x_n\right) = \sum_{j=1}^{n} \left(\int_T \int_{X_j} \omega_{\tau_i}\left(\tilde{x}_j; x_j\right) f_{\tau_j}\left(\tilde{x}_j\right) \mathrm{d}\mu_{\tau_i}\left(\tilde{x}_j; x_j\right) \mathrm{d}\beta_{\tau_i}\right),$$

where for $\tilde{x}_j = x_j$ the weight $\omega_{\tau_i}(x_j; x_j)$ for *experienced* felicity is positive, and for $\tilde{x}_j \neq x_j$ the weight $\omega_{\tau_i}(\tilde{x}_j; x_j)$ for *forgone* or *counterfactual, unexperienced* felicity is non-positive (representing regret), and where the probability measure $\mu_{\tau_i}(\cdot; x_j)$ on X_j expresses the relative importance that τ_i attaches to the experienced felicity of x_j by self j versus the unexperienced, forgone felicity of other $\tilde{x}_j \neq x_j$ by self j.

[14]The space of utility functions is endowed with the Borel σ-algebra of the topology of pointwise convergence.

[15]We need to specify b_{τ_i} explicitly for each type $\tau_i \in T_i$, because u_{τ_i} and β_{τ_i} on their own (i) might sometimes be compatible with several best reply functions (when ties can be broken in several ways), and (ii) might sometimes be compatible with no best reply function at all, when u_{τ_i} has discontinuities, or when for some $j > i$ discontinuities in the best replies b_{τ_j} of types τ_j in the support of β_{τ_i} hinder the expectation of u_{τ_i} with respect to β_{τ_i} from attaining a maximum in X_i (see examples of such situations in Hellwig and Leininger 1987 and Hellwig et al. 1990). Thus, τ_i being measurably associated with $(\beta_{\tau_i}, u_{\tau_i}, b_{\tau_i})$ means in particular that $\beta_{\tau_i}, u_{\tau_i}, b_{\tau_i}$ are mutually compatible.

[16]The space of best-reply functions is endowed with the Borel σ-algebra of the topology of pointwise convergence.

defines a type space with a single state $\tau = (\tau_1, ..., \tau_n)$ where $u_{\tau_i} = u_i$, $\beta_{\tau_i}(\{\tau\}) = 1$ and $b_{\tau_i} = b_i$ for $i = 1, ..., n$.

In the state of the world $\tau = (\tau_1, ..., \tau_n)$ the backward induction path $\hat{x}_\tau = (\hat{x}_{\tau,1}, ..., \hat{x}_{\tau,n})$ is defined inductively by

$$\begin{aligned}
\hat{x}_{\tau,1} &= b_{\tau_1}(\emptyset), \\
\hat{x}_{\tau,2} &= b_{\tau_2}(\hat{x}_{\tau,1}), \\
\hat{x}_{\tau,3} &= b_{\tau_3}(\hat{x}_{\tau,1}, \hat{x}_{\tau,2}), \\
&\vdots \\
\hat{x}_{\tau,n} &= b_{\tau_n}(\hat{x}_{\tau,1}, \hat{x}_{\tau,2}, ..., \hat{x}_{\tau,n-1}).
\end{aligned}$$

Since each type remembers correctly past types and choices, the (expected) well-being levels at τ under backward induction are

$$\begin{aligned}
\hat{u}_\tau &= (\hat{u}_{\tau_1}, ..., \hat{u}_{\tau_i}, ..., \hat{u}_{\tau_n}) \\
&= \Bigg(\int\limits_{T_2 \times ... \times T_n} u_{\tau_1}\left(\hat{x}_{(\tau_1, \tilde{\tau}_2, ... \tilde{\tau}_n)}\right) \mathrm{d}\beta_{\tau_1}(\tilde{\tau}_2, ..., \tilde{\tau}_n), ..., \\
&\int\limits_{T_{i+1} \times ... \times T_n} u_{\tau_i}\left(\hat{x}_{(\tau_1, ... \tau_i, \tilde{\tau}_{i+1}, ... \tilde{\tau}_n)}\right) \mathrm{d}\beta_{\tau_i}(\tilde{\tau}_{i+1}, ..., \tilde{\tau}_n), ..., u_{\tau_n}(\hat{x}_\tau) \Bigg).
\end{aligned}$$

We assume, for simplicity, that the parent knows the true, prevailing state of the world τ, i.e. the parent knows how the beliefs, preferences and best replies of the child are about to evolve.[17] We further assume that if the parent intervenes and induces a path $x \in X$, the parent is open and honest to the child from the very start about the path that the child is henceforth about to follow,[18] which would therefore induce the well-being levels

$$u_{\tau_1}(x), ..., u_{\tau_n}(x).$$

At the state of the world $\tau = (\tau_1, ..., \tau_n) \in T$, a path $x \in X$ that the parent may induce is a *Pareto improvement over backward induction* if $u_{\tau_i}(x) \geq \hat{u}_{\tau_i}$ for $i = 1, ..., n$, and $u_{\tau_j}(x) > \hat{u}_{\tau_j}$ for some $1 \leq j \leq n$.

The state τ represents the utilities and beliefs of the child's selves *after* any information exchange between the parent and the child has already taken place. *Ex ante* information exchange, to the extent that it influences the beliefs of the

[17]See the discussion in section 6 on relaxing this assumption.

[18]In section 6 we will discuss an extension, where even though the parent actually knows the prevailing state of the world $\tau = (\tau_1, ..., \tau_n)$, she can openly and honestly announce a *state-contingent* policy, by which the induced action x_i in period i may depend on the realized type $\tilde{\tau}_i$ in that period.

child, amounts to altering the prevailing state of the world to some other state $\tau' \in T$.

As before, the parent has an average/aggregation function $V : \mathbb{R}^n \to \mathbb{R}$ over the utility levels of the different selves $i = 1, ..., n$, that the parent wants to maximize. In each state of the world $\tau = (\tau_1, ..., \tau_n) \in T$, this induces a utility function of the parent $v_\tau : X \to \mathbb{R}$ over choice paths, defined by

$$v_\tau (x) = V (u_{\tau_1} (x), ..., u_{\tau_n} (x)).$$

3 With misperceptions, Pareto-improving interventions might resiliently lack

For naive or partially naive types, who misperceive their future preferences, there might exist no path that all selves will at least weakly prefer over (their anticipation from) backward induction choices, with strict preference for some self. This is demonstrated in the following example.

Example 2. With the feasible action paths and utility functions specified in example 1 above, consider a type space with two states of the world, $(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)$ and $(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)$, with the beliefs

$$\begin{gathered} \beta_{\dot{\tau}_1} \{(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)\} = 1, \\ \beta_{\dot{\tau}_2} \{(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)\} = 1, \quad \beta_{\ddot{\tau}_2} \{(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)\} = 1, \\ \beta_{\dot{\tau}_3} \{(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)\} = 1, \quad \beta_{\ddot{\tau}_3} \{(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)\} = 1, \end{gathered}$$

and the utility functions

$$\begin{gathered} u_{\dot{\tau}_1} = u_1, \\ u_{\dot{\tau}_2} = u_2, \quad u_{\ddot{\tau}_2} = u_3, \\ u_{\dot{\tau}_3} = u_{\ddot{\tau}_3} = u_3. \end{gathered}$$

The best replies of the types are uniquely determined by these beliefs and utilities, and are computed in the appendix.

In state $(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)$ the selves have the same utility functions as in the previous example, but self 1 naively believes that as of tomorrow the present bias will miraculously disappear ($u_{\ddot{\tau}_2} = u_3$); he moreover wrongly believes that this is commonly known: the state $(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)$, in which $\dot{\tau}_1$ believes, is common knowledge at $(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)$.

Moreover, in state $(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)$ the selves of all three periods, $u_{\dot{\tau}_1}$, $u_{\ddot{\tau}_2}$ and $u_{\ddot{\tau}_3}$, value equally the leisure time at periods 2 and 3. Therefore, by the principle of optimality, the backward induction path $\hat{x}_{(\dot{\tau}_1,\ddot{\tau}_2,\ddot{\tau}_3)}$ already maximizes $u_{\dot{\tau}_1}$ over all possible paths. Since $u_{\dot{\tau}_1}$ is strictly concave, any other path $x \neq \hat{x}_{(\dot{\tau}_1,\ddot{\tau}_2,\ddot{\tau}_3)}$ would actually decrease $u_{\dot{\tau}_1}$, i.e. $u_{\dot{\tau}_1}(x) < \hat{u}_{\dot{\tau}_1}$. In other words, there exists no intervention by the parent which would not decrease the well-being of $\dot{\tau}_1$ in the state $(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)$.

But this means that also in the state $(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)$ there exists no Pareto-improving parental intervention, because in period 1 type $\dot{\tau}_1$ is mistakenly certain there that the state is $(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)$ and expects the backward induction path there $\hat{x}_{(\dot{\tau}_1,\ddot{\tau}_2,\ddot{\tau}_3)}$. However, if the parent attaches equal importance to the well-being of the child in all three periods, i.e. has the same utility v as in example 1, there do exist $x \neq \hat{x}_{(\dot{\tau}_1,\ddot{\tau}_2,\ddot{\tau}_3)}$ with which $v(x) > v\left(\hat{x}_{(\dot{\tau}_1,\dot{\tau}_2,\dot{\tau}_3)}\right)$ (see the details in the appendix). Nevertheless, as long as the parent confines herself to interventions that all selves would consider to be Pareto-improving given their (possibly misguided) beliefs, the parent's hands are tied.

This is not a knife-edge phenomenon. Rather, it is resilient to some perturbations of self 1's beliefs. For example, for $\varepsilon > 0$ small enough (in fact, for $\varepsilon \leq \frac{1}{2}$), even if $\dot{\tau}_1$ ascribes probability ε to the prevailing state of the world $(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)$, i.e.

$$\begin{aligned}
\beta_{\dot{\tau}_1}\{(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)\} &= \varepsilon, & \beta_{\dot{\tau}_1}\{(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)\} &= 1-\varepsilon, \\
\beta_{\dot{\tau}_2}\{(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)\} &= 1, & \beta_{\ddot{\tau}_2}\{(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)\} &= 1, \\
\beta_{\dot{\tau}_3}\{(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)\} &= 1, & \beta_{\ddot{\tau}_3}\{(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)\} &= 1,
\end{aligned}$$

there still do not exist parental interventions which would be Pareto-improving over backward induction. The details are elaborated in the appendix.

3.1 Nudge

Instead of influencing the choice path *given* the misperceptions of the child, the parent can try to intervene *ex ante* with information, by drawing the attention of the child at the state $(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)$ to the actual utilities there $(u_{\dot{\tau}_1}, u_{\dot{\tau}_2}, u_{\dot{\tau}_3})$. That would amount to altering the belief $\beta_{\dot{\tau}_1}$ of the type $\dot{\tau}_1$, by increasing the probability ε that it ascribes to $(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)$.

Such informational, non-coercive interventions are a particular form of a *nudge* (Thaler and Sunstein 2008). In case the nudge is fully successful, $\varepsilon = 1$, the child becomes sophisticated and anticipates correctly his forthcoming preference change.

However, as ε increases, $\dot{\tau}_1$ becomes more disillusioned, and as a result its expected backward-induction well-being $\hat{u}_{\dot{\tau}_1}$ decreases (see the details in the appendix). Thus, even though the nudge does not interfere with the choice x of the child by coercion or incentives, the sobering effect of the nudge causes a backlash to the well-being of the child in period 1, and for no $\varepsilon > 0$ would the nudge induce a Pareto improvement.

4 Minimally disappointing interventions

Situations as in example 2 put the parent in a very frustrating position: due to the child's misperceptions about his future preferences, there is no way for the parent to ameliorate the aggregate well-being of the child without upsetting him at least in one period. This raises the question whether the Pareto-improvement constraint is not too stringent under such misperceptions.

A path x which is Pareto-improving at a state of the world τ vis-à-vis backward induction satisfies, by definition, $u_{\tau_i}(x) - \hat{u}_{\tau_i} \geq 0$ for $i = 1, ..., n$. A natural way to minimally relax this constraint is to require

$$u_{\tau_i}(x) - \hat{u}_{\tau_i} \geq -\delta, \quad i = 1, ..., n, \tag{$\sharp$}$$

for the minimal δ with which a path x satisfying ($\sharp$) exists.

Then, if for this

$$\delta_{\min}(\tau) \equiv \min\left\{\delta : \exists x \ \text{ s.t. } u_{\tau_i}(x) - \hat{u}_{\tau_i} \geq -\delta, \quad i = 1, ..., n\right\},$$

it is the case that

$$\max_{\left\{x:\ \min_{i=1,\ldots,n}\left(u_{\tau_i}(x) - \hat{u}_{\tau_i}\right) \geq -\delta_{\min}(\tau)\right\}} v(x) > v(\hat{x}_\tau),$$

then the parent may find it legitimate to intervene and induce

$$\tilde{x} \in \arg \max_{\left\{x:\ \min_{i=1,\ldots,n}\left(u_{\tau_i}(x) - \hat{u}_{\tau_i}\right) \geq -\delta_{\min}(\tau)\right\}} v(x)$$

in order to augment the aggregate well-being of the child, even though at some period the child will be minimally disappointed, by $\delta_{\min}(\tau)$.

Example 2 (continued). In the state of the world $(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)$ there turns out to be a unique path $\tilde{x}$ satisfying

$$u_{\dot{\tau}_i}(x) - \hat{u}_{\dot{\tau}_i} \geq -\delta_{\min}(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3), \quad i = 1, 2, 3$$

(see the details in the appendix). With this path $\tilde{x}$,

$$u_{\dot{\tau}_1}(\tilde{x}) - \hat{u}_{\dot{\tau}_1} = -\delta_{\min}(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3),$$
$$u_{\dot{\tau}_2}(\tilde{x}) - \hat{u}_{\dot{\tau}_2} = -\delta_{\min}(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3),$$

but

$$u_{\dot{\tau}_3}(\tilde{x}) - \hat{u}_{\dot{\tau}_3} > 2\delta_{\min}(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3).$$

Thus,

$$v(\tilde{x}) > v\left(\hat{x}_{(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)}\right),$$

and, in terms of average well-being across the three periods, which is the parent's perspective, the minimally disappointing intervention is superior to backward induction.

The approach proposed here may actually be applied also when Pareto improving interventions over backward induction *do* exist. With this approach, $\delta_{\min}$ would have the opposite sign and measure the minimal utility enhancement across the periods. The parent would then consider inducing only Pareto improving paths that maximally improve the minimal well-being enhancements across the different periods. In the appendix we demonstrate how this idea is made operative in example 1.

5 Discussion

Parent uncertainty. We assumed, for simplicity, that the parent knows correctly the state of the world, i.e. that the parent knows the child's preferences and beliefs in all periods. We saw that even under such a simplifying assumption, Pareto improving interventions might not be available to the parent. *A fortiori*, a parent who, more realistically, only has some belief about the states of the world, might have an even narrower scope for interventions which would be Pareto improving or minimally disappointing in all the states that she considers possible. How this scope narrows down with the parent's uncertainty remains open for follow-up inquiry.

State-contingent parental interventions. We assumed, again for simplicity, that if the parent induces the child to follow a particular path x, this forthcoming path becomes known to the child, and that very fact might upset the child in some periods. Such upsetting could be potentially attenuated if the parent were allowed to make state-contingent empty promises, like

> 'I know you are certain that tomorrow your present bias will disappear. I am certain that you are too optimistic about this, but if it so happens that I was wrong and you were right, then as of tomorrow I will alter the path so as to make it up for you as much as possible.'

Notice, though, that comes tomorrow, it might be costly or even impossible to verify the time preference of the child, and in such case such a promise would be not only empty (under our assumption that the parent anticipates correctly the future preferences of the child), but also not credible in the first place.

Alternative normative guidelines. When Pareto improving interventions are lacking, we proposed one possible normative constraint, namely minimizing the disappointment of the child across the time periods relative to the benchmark of its (expected) utility under the backward induction path, that he would follow absent of any parental intervention. An alternative normative constraint could be maximizing the minimal overall well-being across the time periods. Or one could even claim that in the absence of Pareto improving interventions, the parent should simply take the lead and induce a path that maximizes the parent's own view of how to balance the considerations of the child's well-being across the time periods, without imposing on herself any further normative constraints.

To conclude, we introduced a type-space framework for beliefs (and beliefs about beliefs, etc.) about one's future preferences, with the implied backward induction choice that each type will make given its belief. We proved that with sophistication, i.e. when the state of the world is common knowledge, with all types agreeing with one another about future preferences and beliefs, then *generically* a parent has room for committing to induce paths which will Pareto-improve well-being across the time periods.

In contrast, with misperceptions about future preferences, such Pareto-improving interventions need not exist, and even informational nudges might be necessarily disillusioning at least in some time period. For such cases we proposed a possible normative alternative to the Pareto criterion, namely minimizing the maximal disappointment that the intervention entails across the time periods.

References

Alós-Ferrer, C., & Ritzberger, K. (2016). Equilibrium existence for large perfect information games. *Journal of Mathematical Economics*, 62, 5-18.

Ericson, K. M., & Laibson, D. (2019). Intertemporal choice. In *Handbook of Behavioral Economics: Applications and Foundations 1* (Vol. 2, pp. 1-67). North-Holland.

Goldman, S. M. (1979). Intertemporally inconsistent preferences and the rate of consumption. *Econometrica,* 47(3), 621-626.

Goldman, S. M. (1980). Consistent plans. *The Review of Economic Studies*, 47(3), 533-537.

Harris, C. (1985). Existence and characterization of perfect equilibrium in games of perfect information. *Econometrica*, 53(3), 613-628.

Hellwig, M., & Leininger, W. (1987). On the existence of subgame-perfect equilibrium in infinite-action games of perfect information. *Journal of Economic Theory*, 43(1), 55-75.

Hellwig, M., Leininger, W., Reny, P. J., & Robson, A. J. (1990). Subgame perfect equilibrium in continuous games of perfect information: an elementary approach to existence and approximation by discrete games. *Journal of Economic Theory*, 52(2), 406-422.

Hirsch, M. W. (1976). *Differential topology*, volume 33 of Graduate texts in mathematics.

Jackson, M. O., & Yariv, L. (2014). Present bias and collective dynamic choice in the lab. *American Economic Review*, 104(12), 4184-4204.

Jackson, M. O., & Yariv, L. (2015). Collective dynamic choice: the necessity of time inconsistency. *American Economic Journal: Microeconomics*, 7(4), 150-78.

Kang, M. (2019). Pareto-improving tax policies under hyperbolic discounting. *International Tax and Public Finance*, 26(3), 618-660.

Kang, M., & Ye, L. S. (2019). Present bias and corporate tax policies. *Journal of Public Economic Theory*, 21(2), 265-290.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2), 443-478.

O'Donoghue, T., & Rabin, M. (1999). Doing it now or later. *American economic review*, 89(1), 103-124.

Phelps, E., & Pollak, R. (1968). On Second-Best National Saving and Game-Equilibrium Growth. *The Review of Economic Studies*, 35(2), 185-199

Pollak, R. A. (1968). Consistent planning. *The Review of Economic Studies*, 35(2), 201-208.

Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *The review of economic studies*, 23(3), 165-180.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press.

Appendix

Example 1

Self 3 has no effective choice but to complete what its preceding selves haven't done, $b_3(x_1, x_2) = 3 - x_1 - x_2$. By backward induction, self 2 solves

$$\begin{aligned}&\max_{x_2} u_2(x_1, x_2, b_3(x_1, x_2)) \\ &= \max_{x_2}\left((24 - x_1^2) + (24 - x_2^2) + \tfrac{1}{2}\left(24 - (3 - x_1 - x_2)^2\right)\right),\end{aligned}$$

leading to

$$b_2(x_1) = 1 - \frac{1}{3}x_1.$$

Accordingly, self 1 solves

$$\begin{aligned}&\max_{x_1} u_1(x_1, b_2(x_1), b_3(x_1, b_2(x_1))) \\ &= \max_{x_1}\left((24 - x_1^2) + \tfrac{1}{2}\left(24 - \left(1 - \tfrac{1}{3}x_1\right)^2\right) + \right. \\ &\left.\tfrac{1}{2}\left(24 - \left(3 - x_1 - \left(1 - \tfrac{1}{3}x_1\right)\right)^2\right)\right),\end{aligned}$$

leading it to choose

$$\hat{x}_1 = \frac{15}{23} = 0.652\,2.$$

Consequently,

$$\hat{x}_2 = b_2(\hat{x}_1) = \frac{18}{23} = 0.782\,6,$$

and

$$\hat{x}_3 = b_3(\hat{x}_1, \hat{x}_2) = \frac{36}{23} = 1.\,565\,2.$$

Altogether, the backward induction path is $\hat{x} = \left(\frac{15}{23}, \frac{18}{23}, \frac{36}{23}\right)$. The corresponding utilities are

$$\begin{aligned}\hat{u}_1 &= u_1(\hat{x}) = 46.\,044, \\ \hat{u}_2 &= u_2(\hat{x}) = 57.\,737, \\ \hat{u}_3 &= u_3(\hat{x}) = 68.\,512,\end{aligned} \tag{$*$}$$

whose average is

$$V(\hat{u}) = v(\hat{x}) = 57.431. \tag{$\clubsuit$}$$

The parent's problem

With no normative constraints on interventions, the parent would maximize the average utility of the three selves

$$v(x_1, x_2, x_3) = \frac{1}{3}\sum_{i=1}^{3} u_i(x_1, x_2, x_3) = \left(24 - x_1^2\right) + \frac{5}{6}\left(24 - x_2^2\right) + \frac{2}{3}\left(24 - x_3^2\right),$$

subject to the physical constraint $x_1 + x_2 + x_3 = 3$. This yields

$$\bar{x} \equiv (\bar{x}_1, \bar{x}_2, \bar{x}_3) = \left(\frac{30}{37}, \frac{36}{37}, \frac{45}{37}\right) = (0.811, 0.973, 1.216).$$

The utility levels are

$$\begin{aligned} u_1(\bar{x}) &= 46.13, \\ u_2(\bar{x}) &= 57.656, , \\ u_3(\bar{x}) &= 68.917. \end{aligned} \tag{$**$}$$

whose average is

$$v(\bar{x}) = 57.568, \tag{♣♣}$$

is higher than without intervention, (♣).

However, comparing ($**$) to ($*$) reveals that $\bar{x}$ is not a Pareto improvement relative to the backward induction path without intervention $\hat{x}$, because $u_2(\bar{x}) < \hat{u}_2$.

Pareto-improving interventions

If the parent restricts herself to interventions that are Pareto-improving relative the backward induction path $\hat{x}$, the parent solves

$$\max_{x_1+x_2+x_3=3} v(x_1, x_2, x_3), \quad \text{s.t.} \quad u_i(x_1, x_2, x_3) \geq \hat{u}_i, \quad i = 1, 2, 3,$$

yielding

$$\breve{x} \equiv (\breve{x}_1, \breve{x}_2, \breve{x}_3) = (0.780, 0.828, 1.392),$$

with utility levels

$$\begin{aligned} u_1(\breve{x}) &= 46.080, \\ u_2(\breve{x}) &= 57.737, \\ u_3(\breve{x}) &= 68.768, \end{aligned} \tag{$***$}$$

with lower average

$$v(\breve{x}) = 57.528 < v(\bar{x}). \qquad (\clubsuit\clubsuit\clubsuit)$$

In particular, along the choice path $\breve{x}$ self 1 does more of the chore relative to the backward induction path $\hat{x}$, but (comparing $(*)$ with $(***)$) self 1 it is more than happy to do so knowing that self 2 will also work harder; self 2 is just indifferent working harder given that self 1 works harder; and self 3 is happier that less of the largest share of the chore is left for it to complete.

Maximin intervention

If the parent restricts herself further, to an intervention that maximizes the minimal utility enhancement across the three selves relative to the no-intervention backward induction path $\hat{x}$,

$$\max_{x_1+x_2+x_3=1} \min_{i=1,2,3} \left[u_i(x_1, x_2, x_3) - \hat{u}_i\right].$$

The maximin is attained at

$$\tilde{x} \equiv (\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) = (0.739, 0.784, 1.477),$$

with utility levels

$$\begin{aligned} u_1(\tilde{x}) &= 46.056, \\ u^{\hat{\imath}}_{\tau_2}(\tilde{x}) &= 57.748, \\ u^{\hat{\imath}}_{\tau_3}(\tilde{x}) &= 68.657 \end{aligned} \qquad (****)$$

accruing a utility increment of 0.0115 to *both* self 1 and self 2 relative to the no-intervention backward induction path, even if with an average utility across the selves

$$v(\tilde{x}) = 57.487, \qquad (\clubsuit\clubsuit\clubsuit\clubsuit)$$

which is lower than $(\clubsuit\clubsuit\clubsuit)$. Nevertheless, comparing with $(\clubsuit)$ the parent realizes that $v(\tilde{x}) > v(\hat{x})$, so she judges the maximin intervention as preferable to no intervention at all.

Example 2

Both $\dot{\tau}_3$ and $\ddot{\tau}_3$ have no choice but to complete the part of the chore that the previous selves haven't completed,

$$b_{\dot{\tau}_3}(x_1, x_2) = b_{\ddot{\tau}_3}(x_1, x_2) = 3 - x_1 - x_2.$$

As for self 2, type $\ddot{\tau}_2$, who has no present bias and the same utility function u_3 as that of $\ddot{\tau}_3$, will divide the remaining chore equally among them,

$$b_{\ddot{\tau}_2}(x_1) = \frac{3 - x_1}{2},$$

whereas type $\dot{\tau}_2$, who does have a present bias with the utility function u_2 will, like self 2 in example 1, choose only

$$b_{\dot{\tau}_2}(x_1) = 1 - \frac{1}{3}x_1.$$

The naive type $\dot{\tau}_1$, who is certain that the subsequent types are $\ddot{\tau}_2$,$\ddot{\tau}_3$ will solve

$$\begin{aligned}
&\max_{x_1} u_1\left(x_1, b_{\ddot{\tau}_2}(x_1), b_{\ddot{\tau}_3}(x_1, b_{\ddot{\tau}_2}(x_1))\right) \\
&= \max_{x_1}\left(\left(24 - x_1^2\right) + \frac{1}{2}\left(24 - \left(\frac{3 - x_1}{2}\right)^2\right) + \frac{1}{2}\left(24 - \left(\frac{3 - x_1}{2}\right)^2\right)\right),
\end{aligned}$$

leading him to choose

$$b_{\dot{\tau}_1}(\emptyset) = \hat{x}_{\dot{\tau}_1} = \frac{3}{5},$$

believing that $\ddot{\tau}_2$ and $\ddot{\tau}_3$ will divide the remaining chore equally, each choosing $\frac{6}{5}$. However, at the state $(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)$, type $\dot{\tau}_2$ will actually choose

$$\hat{x}_{\dot{\tau}_2} = b_{\dot{\tau}_2}(\hat{x}_{\dot{\tau}_1}) = 1 - \frac{1}{3} \cdot \frac{3}{5} = \frac{4}{5},$$

leaving

$$\hat{x}_{\dot{\tau}_3} = \frac{8}{5}$$

of the chore to $\dot{\tau}_3$. At $(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)$ the (expected) utilities will therefore be

$$\begin{aligned}
\hat{u}_{\dot{\tau}_1} &= u_{\dot{\tau}_1}\left(\frac{3}{5}, \frac{6}{5}, \frac{6}{5}\right) = 46.2, \\
\hat{u}_{\dot{\tau}_2} &= u_{\dot{\tau}_2}\left(\frac{3}{5}, \frac{4}{5}, \frac{8}{5}\right) = 57.72, \qquad (*****) \\
\hat{u}_{\dot{\tau}_3} &= u_{\dot{\tau}_3}\left(\frac{3}{5}, \frac{4}{5}, \frac{8}{5}\right) = 68.44.
\end{aligned}$$

Thus, with no intervention by the parent, the resulting average utility at $(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)$ will then be

$$V\left(\hat{u}_{\dot{\tau}_1}, \hat{u}_{\dot{\tau}_2}, \hat{u}_{\dot{\tau}_3}\right) = \frac{\hat{u}_{\dot{\tau}_1} + \hat{u}_{\dot{\tau}_2} + \hat{u}_{\dot{\tau}_3}}{3} = 57.453. \qquad (\clubsuit\clubsuit\clubsuit\clubsuit\clubsuit)$$

There exists no Pareto improving intervention

At state $(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)$, no intervention can be Pareto improving over the backward induction utility profile $(\hat{u}_{\dot{\tau}_1}, \hat{u}_{\dot{\tau}_2}, \hat{u}_{\dot{\tau}_3})$, because $\dot{\tau}_1$ believes that the state is $(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)$ in which there is (common knowledge of) time consistency of preferences, and therefore by the principle of optimality in dynamic programming any alteration of the action profile $\left(\frac{3}{5}, \frac{6}{5}, \frac{6}{5}\right)$ chosen sequentially by $\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3$ would be in particular unbeneficial from the perspective of $\dot{\tau}_1$. Moreover, since u_1 is strictly concave, any such alteration would actually decrease u_1.

The absence of Pareto improvements does not hinge on the fact that $\dot{\tau}_1$ is (wrongly) certain of (common knowledge of) the selves' time consistency. To see this, amend the above example so that type $\dot{\tau}_1$ does assign probability $\varepsilon > 0$ that there is time inconsistency. Thus:

$$\begin{aligned}
&\beta_{\dot{\tau}_1}\{(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)\} = \varepsilon, \quad \beta_{\dot{\tau}_1}\{(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)\} = 1-\varepsilon,\\
&\beta_{\dot{\tau}_2}\{(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)\} = 1, \quad \beta_{\ddot{\tau}_2}\{(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)\} = 1,\\
&\beta_{\dot{\tau}_3}\{(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)\} = 1, \quad \beta_{\ddot{\tau}_3}\{(\dot{\tau}_1, \ddot{\tau}_2, \ddot{\tau}_3)\} = 1.
\end{aligned}$$

The maximization problem of $\dot{\tau}_1$ will now be

$$\begin{aligned}
&\max_{x_1}\\
&(\varepsilon u_{\dot{\tau}_1}(x_1, b_{\dot{\tau}_2}(x_1), b_{\dot{\tau}_3}(x_1, b_{\dot{\tau}_2}(x_1))) +\\
&(1-\varepsilon) u_{\dot{\tau}_1}(x_1, b_{\ddot{\tau}_2}(x_1), b_{\ddot{\tau}_3}(x_1, b_{\ddot{\tau}_2}(x_1)))) =\\
&\max_{x_1}\\
&\left(\begin{array}{l} \varepsilon\left((24 - x_1^2) + \frac{1}{2}\left(24 - \left(1 - \frac{1}{3}x_1\right)^2\right) + \frac{1}{2}\left(24 - \left(3 - x_1 - \left(1 - \frac{1}{3}x_1\right)\right)^2\right)\right) \\ +(1-\varepsilon)\left((24 - x_1^2) + \frac{1}{2}\left(24 - \left(\frac{3-x_1}{2}\right)^2\right) + \frac{1}{2}\left(24 - \left(\frac{3-x_1}{2}\right)^2\right)\right) \end{array}\right),
\end{aligned}$$

leading $\dot{\tau}_1$ to choose

$$\hat{x}_{\dot{\tau}_1} = \frac{3\varepsilon + 27}{\varepsilon + 45},$$

and, subsequently,

$$\begin{aligned}
\hat{x}_{\dot{\tau}_2} &= b_{\dot{\tau}_2}(\hat{x}_{\dot{\tau}_1}) = 1 - \frac{1}{3}\frac{3\varepsilon + 27}{\varepsilon + 45} = \frac{36}{\varepsilon + 45},\\
\hat{x}_{\ddot{\tau}_2} &= b_{\ddot{\tau}_2}(\hat{x}_{\dot{\tau}_1}) = \frac{3 - \frac{3\varepsilon+27}{\varepsilon+45}}{2} = \frac{54}{\varepsilon + 45},
\end{aligned}$$

leaving to the last self

$$\begin{aligned}
\hat{x}_{\dot{\tau}_3} &= \frac{72}{\varepsilon + 45},\\
\hat{x}_{\ddot{\tau}_3} &= \frac{54}{\varepsilon + 45}.
\end{aligned}$$

Accordingly, the backward induction expected utilities at $(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)$ will be

$$\begin{aligned}
&\hat{u}_{\dot{\tau}_1} = \varepsilon u_{\dot{\tau}_1}\left(\tfrac{3\varepsilon+27}{\varepsilon+45}, \tfrac{36}{\varepsilon+45}, \tfrac{72}{\varepsilon+45}\right) + \\
&(1-\varepsilon)\, u_{\dot{\tau}_1}\left(\tfrac{3\varepsilon+27}{\varepsilon+45}, \tfrac{54}{\varepsilon+45}, \tfrac{54}{\varepsilon+45}\right) = \tfrac{39\varepsilon+2079}{\varepsilon+45}, \\
&\hat{u}_{\dot{\tau}_2} = u_{\dot{\tau}_2}\left(\tfrac{3\varepsilon+27}{\varepsilon+45}, \tfrac{36}{\varepsilon+45}, \tfrac{72}{\varepsilon+45}\right) = \tfrac{3\left(17\varepsilon^2+1746\varepsilon+38\,961\right)}{(\varepsilon+45)^2}, \\
&\hat{u}_{\dot{\tau}_2} = u_{\dot{\tau}_2}\left(\tfrac{3\varepsilon+27}{\varepsilon+45}, \tfrac{36}{\varepsilon+45}, \tfrac{72}{\varepsilon+45}\right) = \tfrac{9\left(7\varepsilon^2+702\varepsilon+15\,399\right)}{(\varepsilon+45)^2}.
\end{aligned}$$

Then one can verify that for $\varepsilon \leq \frac{1}{2}$ there exist no Pareto improvements $(x_1, x_2, 3 - x_1 - x_2)$ over this utility tuple.

Nudge One can interpret an increase in ε as the result of an *ex ante* nudge, that sobers up self 1 at least partially, and makes it consider the possibility that its present bias need not disappear tomorrow (type $\ddot{\tau}_2$ has no present bias), but might rather persist (type $\dot{\tau}_2$) with probability ε.

Such a nudge, though, will *decrease* the well-being $\hat{u}_{\dot{\tau}_1}$ of self 1 at the no-intervention backward induction path, because

$$\frac{d\hat{u}_{\dot{\tau}_1}}{d\varepsilon} = -\frac{324}{(\varepsilon+45)^2} < 0.$$

Thus, even such an informational intervention, involving no coercion or incentives, is not Pareto improving.

Minimax intervention

With the lack of Pareto-improvements, it may still be the case, though, that by disappointing some of the selves while cheering up others the parent can improve upon the average utility (♣♣♣♣♣) at $(\dot{\tau}_1, \dot{\tau}_2, \dot{\tau}_3)$. One possibility to use the same criterion as before, that amounts in this case to choosing a path that will minimize the maximal disappointment vis-a-vis $(*\,*\,*\,*\,*)$ across the 3 selves

$$\min_{x_1+x_2+x_3=3} \max_{i=1,2,3} \left[\hat{u}_{\dot{\tau}_i} - u_{\dot{\tau}_i}(x_1, x_2, x_3)\right].$$

The solution is $\left(\tilde{x}_{\dot{\tau}_1}, \tilde{x}_{\dot{\tau}_2}, \tilde{x}_{\dot{\tau}_3}\right) = (0.673\,4,\ 0.979\,8, 1.\,346\,8)$, with which $\dot{\tau}_1$ and $\dot{\tau}_2$ lose 0.04 04 of their utility, each, but $\dot{\tau}_3$ gains 0.333, overall leading to the higher average utility

$$v\left(\tilde{x}_{\dot{\tau}_1}, \tilde{x}_{\dot{\tau}_2}, \tilde{x}_{\dot{\tau}_3}\right) = 57.\,537 > v\left(\hat{x}_{\dot{\tau}_1}, \hat{x}_{\dot{\tau}_2}, \hat{x}_{\dot{\tau}_3}\right)$$

as comparing with (♣♣♣♣♣) reveals. Thus, from the parent's perspective this is a worthwhile intervention.

This intervention is in some sense even more 'benign' than the informational nudge considered above, because it makes sure to smooth as much as possible the sobering pain of disillusionment across the different periods, rather than weighing its entire toll on the child in the first period only.

Proof of the Theorem

For simplicity, we first present the proof for the case where the choice set of each self is one-dimensional, $X_i \subset \mathbb{R}^{k_i}$ with $k_i = 1$, $i = 1, ..., n$, and then elaborate on how to read the same proof for the case of any finite $k_i \geq 1$.

If $\hat{x} = (\hat{x}_1, ..., \hat{x}_n)$ is an interior subgame-perfect equilibrium path, then at $\hat{x}$ the following n first-order conditions obtain:

$$\frac{\partial u_n}{\partial x_n} = 0,$$

$$\frac{\partial u_{n-1}}{\partial x_{n-1}} + \frac{\partial u_{n-1}}{\partial x_n}\frac{\partial b_n}{\partial x_{n-1}} = 0,$$

$$\frac{\partial u_{n-2}}{\partial x_{n-2}} + \frac{\partial u_{n-2}}{\partial x_{n-1}}\frac{\partial b_{n-1}}{\partial x_{n-2}} + \frac{\partial u_{n-2}}{\partial x_n}\left(\frac{\partial b_n}{\partial x_{n-2}} + \frac{\partial b_n}{\partial x_{n-1}}\frac{\partial b_{n-1}}{\partial x_{n-2}}\right) = 0,$$

$$\vdots$$

To simplify notation in the sequel, define the matrix of direct and indirect effects

$$h = \begin{pmatrix} 1 & & \cdots & & & \\ 0 & 1 & & & & \vdots \\ \vdots & & \ddots & & & \\ 0 & & 0 & 1 & \frac{\partial b_{n-1}}{\partial x_{n-2}} & \left(\frac{\partial b_n}{\partial x_{n-2}} + \frac{\partial b_n}{\partial x_{n-1}}\frac{\partial b_{n-1}}{\partial x_{n-2}}\right) \\ 0 & \cdots & & 0 & 1 & \frac{\partial b_n}{\partial x_{n-1}} \\ 0 & & \cdots & & 0 & 1 \end{pmatrix},$$

so that the above system becomes, at $\hat{x}$,

$$F_n(x; u) \equiv \frac{\partial u_n}{\partial x_n} = 0, \qquad (BI_{\text{n}})$$

$$F_{n-1}(x; u) \equiv \frac{\partial u_{n-1}}{\partial x_{n-1}} + \frac{\partial u_{n-1}}{\partial x_n}h_{n-1,n} = 0, \qquad (BI_{\text{n-1}})$$

$$F_{n-2}(x; u) \equiv \frac{\partial u_{n-2}}{\partial x_{n-2}} + \frac{\partial u_{n-2}}{\partial x_{n-1}}h_{n-2,n-1} + \frac{\partial u_{n-2}}{\partial x_n}h_{n-2,n} = 0. \qquad (BI_{\text{n-2}})$$

$$\vdots$$

If $\hat{x}$ is also Pareto optimal, then there exists $\hat{\theta} = \left(\hat{\theta}_1, ..., \hat{\theta}_n\right)$ such that at $\left(\hat{x}, \hat{\theta}\right)$

$$G_1\left(x, \theta; u\right) \equiv \sum_{i=1}^{n} \theta_i \frac{\partial u_i}{\partial x_1} = 0, \tag{PO_1}$$

$$\vdots$$

$$G_n\left(x, \theta; u\right) \equiv \sum_{i=1}^{n} \theta_i \frac{\partial u_i}{\partial x_n} = 0, \tag{PO_n}$$

and

$$G_{n+1}\left(\theta\right) = \sum_{i=1}^{n} \theta_i^2 - 1 = 0. \tag{$PO_{\mathrm{n+1}}$}$$

Indeed, if there doesn't exist $\left(\hat{\theta}_1, ..., \hat{\theta}_n\right)$ satisfying (PO_{n+1}) such that $(PO_1), ...,$ (PO_n) hold, the Jacobian Du has full row rank at $\hat{x}$, and one can find a direction $\begin{pmatrix} \delta_1 \\ \vdots \\ \delta_n \end{pmatrix}$ such that $Du \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_n \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$, so that perturbing $\hat{x}_i$ in the direction δ_i for $i = 1, ..., n$ improves all of $u_1, ...u_n$, thus constituting a Pareto improvement.

We will distinguish two families of cases: (i) when there is a minimal index $a \geq 1$ and a maximal index $b > a$ such that $\hat{\theta}_a \neq 0$ and $\hat{\theta}_b \neq 0$; (ii) when there is a unique index $1 \leq c \leq n$ for which $\hat{\theta}_c \neq 0$ while $\hat{\theta}_j = 0\ \forall j \neq c$.

For every case of the first family, we will show that for every perturbation direction $(\pi_1, ..., \pi_n, p_1, ..., p_n, p_{n+1})$ for $(F_1, ..., F_n, G_1, ...G_n, G_{n+1})$ there exists

$$Y = \begin{pmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & & \vdots \\ y_{n1} & \cdots & y_{nn} \end{pmatrix} \in \mathbb{R}^{n \times n},$$

such that, with the path of utility tuples

$$u^t = \begin{pmatrix} u_1^t \\ \vdots \\ u_n^t \end{pmatrix} \equiv u + tY \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

defined for $t \in [-\varepsilon, \varepsilon]$ for some $\varepsilon > 0$, and with

$$G_{n+1}^t = G_{n+1} + tp_{n+1},$$

we get at $t = 0$

$$\frac{\partial F_1\left(\hat{x}; u^0\right)}{\partial t} = \pi_1,$$
$$\vdots$$
$$\frac{\partial F_n\left(\hat{x}; u^0\right)}{\partial t} = \pi_n,$$
$$\frac{\partial G_1\left(\hat{x}, \hat{\theta}; u^0\right)}{\partial t} = p_1,$$
$$\vdots$$
$$\frac{\partial G_n\left(\hat{x}, \hat{\theta}; u^0\right)}{\partial t} = p_n,$$
$$\frac{\partial G^0_{n+1}\left(\hat{\theta}\right)}{\partial t} = p_{n+1}.$$

The last equation is immediate by the definition of G^t_{n+1}. For the first $2n$ equations to hold, choose

$$Y =$$

$$\begin{pmatrix}
\pi_1 & 0 & \cdots & 0 & & & \cdots & 0 \\
0 & \ddots & & 0 & & & & 0 \\
\vdots & & \ddots & \vdots & & & & \vdots \\
\frac{p_1-\hat{\theta}_1\pi_1}{\hat{\theta}_a} & \cdots & \cdots & \pi_a - \sum\limits_{i=a+1}^{n} \frac{p_i-\hat{\theta}_i\pi_i}{\hat{\theta}_a} h_{a,i} & \cdots & \frac{p_b-\hat{\theta}_b\pi_b}{\hat{\theta}_a} & \cdots & \frac{p_n-\hat{\theta}_n\pi_n}{\hat{\theta}_a} \\
\vdots & & & \vdots & \ddots & \vdots & & \vdots \\
0 & \cdots & \cdots & \frac{p_a-\hat{\theta}_a\left(\pi_a - \sum\limits_{i=a+1}^{n} \frac{p_i-\hat{\theta}_i\pi_i}{\hat{\theta}_a} h_{a,i}\right)}{\hat{\theta}_b} & 0 & \pi_b & & 0 \\
\vdots & & & \vdots & & & \ddots & \vdots \\
0 & \cdots & \cdots & 0 & & 0 & & \pi_n
\end{pmatrix}.$$

In words,

1. for $i \neq a$, set $y_{ii} = \pi_i$, thus perturbing F_i at $t = 0$ by π_i for $i \neq a$,

2. for $i \neq a$, (1) per.turbs G_i at $t = 0$ by $\hat{\theta}_i\pi_i$, so set $y_{ai} = \frac{p_i-\hat{\theta}_i\pi_i}{\hat{\theta}_a}$ in order to eventually end up perturbing G_i at $t = 0$ by p_i,

3. together over all $i \neq a$, (2) upsets F_a at $t=0$ by $\sum_{i>a} \frac{p_i - \hat{\theta}_i \pi_i}{\hat{\theta}_a} h_{a,i}$ (only, $h_{a,i}$ itself was not perturbed, because it depends on partial cross-derivatives of u_r w.r.t. x_s for $r > s > a$, and these were not perturbed) and, in order to eventually perturb F_i at $t=0$ by π_a, set $y_{aa} = \pi_a - \sum_{i>a} \frac{p_i - \hat{\theta}_i \pi_i}{\hat{\theta}_a} h_{a,i}$ and,

4. finally, (3) upsets G_a at $t=0$ by $\hat{\theta}_a \left(\pi_a - \sum_{i>a} \frac{p_i - \hat{\theta}_i \pi_i}{\hat{\theta}_a} h_{a,i} \right)$, so in order to eventually perturb G_a at $t=0$ by p_a, set

$$y_{ba} = \frac{p_a - \hat{\theta}_a \left(\pi_a - \sum_{i>a} \frac{p_i - \hat{\theta}_i \pi_i}{\hat{\theta}_a} h_{a,i} \right)}{\hat{\theta}_b},$$

which does not upset F_b, which only depends on partial derivatives of u_{ib} w.r.t. x_s for $s \geq b$, whereas $a < b$), while

5. all the remaining entries of Y are zero.

Therefore, by the transversality theorem there exists an open and dense subset $\mathcal{U}_{a,b} \subseteq \mathcal{U}$ of utility profiles u for which at any solution $\left(\hat{x}, \hat{\theta}\right)$ of

$$(BI_1), ..., (BI_n), (PO_1), ..., (PO_n), (PO_{n+1}),$$

for which $\arg\min_i \left(\hat{\theta}_i \neq 0\right) = a$ and $\arg\max_i \left(\hat{\theta}_i \neq 0\right) = b$, the $(2n+1) \times 2n$ matrix

$$\begin{pmatrix} \frac{\partial F_1(\hat{x};u)}{\partial x_1} & \cdots & \frac{\partial F_1(\hat{x};u)}{\partial x_n} & \frac{\partial F_1(\hat{x};u)}{\partial \theta_1} & \cdots & \frac{\partial F_1(\hat{x};u)}{\partial \theta_n} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial F_n(\hat{x};u)}{\partial x_1} & \cdots & \frac{\partial F_n(\hat{x};u)}{\partial x_n} & \frac{\partial F_n(\hat{x};u)}{\partial \theta_1} & \cdots & \frac{\partial F_n(\hat{x};u)}{\partial \theta_n} \\ \frac{\partial G_1\left(\hat{x},\hat{\theta};u\right)}{\partial x_1} & \cdots & \frac{\partial G_1\left(\hat{x},\hat{\theta};u\right)}{\partial x_n} & \frac{\partial G_1\left(\hat{x},\hat{\theta};u\right)}{\partial \theta_1} & \cdots & \frac{\partial G_1\left(\hat{x},\hat{\theta};u\right)}{\partial \theta_n} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial G_n\left(\hat{x},\hat{\theta};u\right)}{\partial x_1} & \cdots & \frac{\partial G_n\left(\hat{x},\hat{\theta};u\right)}{\partial x_n} & \frac{\partial G_n\left(\hat{x},\hat{\theta};u\right)}{\partial \theta_1} & \cdots & \frac{\partial G_n\left(\hat{x},\hat{\theta};u\right)}{\partial \theta_n} \\ \frac{\partial G_{n+1}\left(\hat{\theta}\right)}{\partial x_1} & \cdots & \frac{\partial G_{n+1}\left(\hat{\theta}\right)}{\partial x_n} & \frac{\partial G_{n+1}\left(\hat{\theta}\right)}{\partial \theta_1} & \cdots & \frac{\partial G_{n+1}\left(\hat{\theta}\right)}{\partial \theta_n} \end{pmatrix} \quad (\bigstar)$$

has full row rank.[19]

As for cases of the second family, where at a solution $\left(\hat{x}, \hat{\theta}\right)$, $\hat{\theta}_c$ is the only non-zero entry in $\hat{\theta}$, $(PO_1), ..., (PO_n)$ amount to $\nabla u_c = \left(\frac{\partial u_c}{\partial x_1}, ..., \frac{\partial u_c}{\partial x_n}\right) = 0$, which

[19]Hirsch (1976), p. 74, theorem 2.1 (b), with $M = \mathbb{R}^{2n}$, $N = \mathbb{R}^{2n+1}$, $L = X \times [0,1]^n$ and $A = \{0\}$. For this case the theorem then states that $\pitchfork^2_{X \times [0,1]^n} \left(\mathbb{R}^{2n}, \mathbb{R}^{2n+1}; \{0\}\right)$ is open and dense within $C^2 \left(\mathbb{R}^{2n}, \mathbb{R}^{2n+1}\right)$.

turns (BI_c) to an identity, $0=0$. In such a case renounce therefore the expression F_c, but also read the remaining $2n$ expressions $(F_j)_{j\neq c}, (G_i)_{i=1}^{n+1}$ as functions of the $n+1$ variables $x_1,...,x_n,\theta_c$ only, i.e. with $\theta_j \equiv 0$ for $j\neq c$. Then for every corresponding rates of perturbation $\left((\pi_j)_{j\neq c}, (p_i)_{i=1}^{n+1}\right)$, if we choose

$$Y=\begin{pmatrix} \pi_1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \pi_j & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ \frac{p_1}{\hat{\theta}_c} & \cdots & \frac{p_j}{\hat{\theta}_c} & \cdots & \frac{p_c}{\hat{\theta}_c} & & \frac{p_n}{\hat{\theta}_c} \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & \pi_n \end{pmatrix},$$

then at $t=0$ we have $\frac{\partial F_j\left(\hat{x};u^0\right)}{\partial t}=\pi_j$ for $j\neq c$, $\frac{\partial G_i\left(\hat{x},\hat{\theta}_c;u^0\right)}{\partial t}=p_i$ for $i=1,...,n$, and $\frac{\partial G^0_{n+1}\left(\hat{\theta}_c\right)}{\partial t}=p_{n+1}$.

Therefore, by the transversality theorem there exists an open and dense subset $\mathcal{U}_c\subseteq\mathcal{U}$ of utility profiles u for which at any solution $\left(\hat{x},\hat{\theta}\right)$ of

$$(BI_1),...,(BI_n),(PO_1),...,(PO_n),(PO_{n+1}),$$

for which $\hat{\theta}_c\neq 0$ and $\hat{\theta}_j=0$ for $j\neq c$, the $2n\times(n+1)$ matrix

$$\begin{pmatrix} \frac{\partial F_1(\hat{x};u)}{\partial x_1} & \cdots & \frac{\partial F_1(\hat{x};u)}{\partial x_n} & \frac{\partial F_1(\hat{x};u)}{\partial \theta_c} \\ \vdots & & \vdots & \vdots \\ \frac{\partial F_{c-1}(\hat{x};u)}{\partial x_1} & \cdots & \frac{\partial F_{c-1}(\hat{x};u)}{\partial x_n} & \frac{\partial F_{c-1}(\hat{x};u)}{\partial \theta_c} \\ \frac{\partial F_{c+1}(\hat{x};u)}{\partial x_1} & \cdots & \frac{\partial F_{c+}(\hat{x};u)}{\partial x_n} & \frac{\partial F_{c+}(\hat{x};u)}{\partial \theta_c} \\ \vdots & & \vdots & \vdots \\ \frac{\partial F_n(\hat{x};u)}{\partial x_1} & \cdots & \frac{\partial F_n(\hat{x};u)}{\partial x_n} & \frac{\partial F_n(\hat{x};u)}{\partial \theta_c} \\ \frac{\partial G_1\left(\hat{x},\hat{\theta}_c;u\right)}{\partial x_1} & \cdots & \frac{\partial G_1\left(\hat{x},\hat{\theta}_c;u\right)}{\partial x_n} & \frac{\partial G_1\left(\hat{x},\hat{\theta}_c;u\right)}{\partial \theta_c} \\ \vdots & & \vdots & \vdots \\ \frac{\partial G_n\left(\hat{x},\hat{\theta}_c;u\right)}{\partial x_1} & \cdots & \frac{\partial G_n\left(\hat{x},\hat{\theta}_c;u\right)}{\partial x_n} & \frac{\partial G_n\left(\hat{x},\hat{\theta}_c;u\right)}{\partial \theta_c} \\ \frac{\partial G_{n+1}\left(\hat{\theta}_c\right)}{\partial x_1} & \cdots & \frac{\partial G_{n+1}\left(\hat{\theta}_c\right)}{\partial x_n} & \frac{\partial G_{n+1}\left(\hat{\theta}_c\right)}{\partial \theta_c} \end{pmatrix} \qquad (\bigstar\bigstar)$$

has full row rank.

The intersection of the finitely many open and dense sets

$$\mathcal{U}_0=\left(\bigcap_{1\leq a<b\leq n}\mathcal{U}_{a,b}\right)\cap\left(\bigcap_{c=1}^{n}\mathcal{U}_c\right)$$

is open and dense. For every utility tuple $u \in \mathcal{U}_0$, at every solution $\left(\hat{x}, \hat{\theta}\right)$ of $(BI_1), ..., (BI_n), (PO_1), ..., (PO_n), (PO_{n+1})$, either ($\bigstar$) or ($\bigstar\bigstar$) has full row rank, which is impossible because both ($\bigstar$) and ($\bigstar\bigstar$) have more rows than columns. Hence for generic u (namely for $u \in \mathcal{U}_0$), no interior backward induction SPE path $\hat{x}$ is Pareto optimal.

When choice variables are multi-dimensional, $x_i = (x_{i,1}, ..., x_{i,k_i})$, $i = 1, ..., n$, the same proof applies verbatim with the following caveats: $\frac{\partial u_i}{\partial x_j} = \left(\frac{\partial u_i}{\partial x_{j,1}}, \dots \frac{\partial u_i}{\partial x_{j,k_j}}\right)$, $\frac{\partial b_i}{\partial x_j} = \begin{pmatrix} \frac{\partial b_{i,1}}{\partial x_{j,1}} & \cdots & \frac{\partial b_{i,1}}{\partial x_{j,k_j}} \\ \vdots & & \vdots \\ \frac{\partial b_{i,k_i}}{\partial x_{j,1}} & \cdots & \frac{\partial b_{i,k_i}}{\partial x_{j,k_j}} \end{pmatrix}$, $F_i = (F_{i,1}, ..., F_{i,k_i})$, $\frac{\partial F_i}{\partial x_j} = \begin{pmatrix} \frac{\partial F_{i,1}}{\partial x_{j,1}} & \cdots & \frac{\partial F_{i,1}}{\partial x_{j,k_j}} \\ \vdots & & \vdots \\ \frac{\partial F_{i,k_i}}{\partial x_{j,1}} & \cdots & \frac{\partial F_{i,k_i}}{\partial x_{j,k_j}} \end{pmatrix}$, $\pi_i = (\pi_{i,1}, ..., \pi_{i,k_i})$, $G_i = (G_{i,1}, ..., G_{i,k_i})$, $\frac{\partial G_i}{\partial x_j} = \begin{pmatrix} \frac{\partial G_{i,1}}{\partial x_{j,1}} & \cdots & \frac{\partial G_{i,1}}{\partial x_{j,k_j}} \\ \vdots & & \vdots \\ \frac{\partial G_{i,k_i}}{\partial x_{j,1}} & \cdots & \frac{\partial G_{i,k_i}}{\partial x_{j,k_j}} \end{pmatrix}$, $p_i = (p_{i,1}, ..., p_{i,k_i})$, $\delta_i = (\delta_{i,1}, ..., \delta_{i,k_i})$; θ_i remains a scalar. ■