A new characterization of endogeny

Tibor Mach [*] Anja Sturm [†] Jan M. Swart [*]

October 1, 2018

Abstract

Aldous and Bandyopadhyay have shown that each solution to a recursive distributional equation (RDE) gives rise to a recursive tree process (RTP), which is a sort of Markov chain in which time has a tree-like structure and in which the state of each vertex is a random function of its descendants. If the state at the root is measurable with respect to the sigma field generated by the random functions attached to all vertices, then the RTP is said to be endogenous. For RTPs defined by continuous maps, Aldous and Bandyopadhyay showed that endogeny is equivalent to bivariate uniqueness, and they asked if the continuity hypothesis can be removed. We introduce a higher-level RDE that through its n-th moment measures contains all n-variate RDEs. We show that this higher-level RDE has minimal and maximal fixed points with respect to the convex order, and that these coincide if and only if the corresponding RTP is endogenous. As a side result, this allows us to answer the question of Aldous and Bandyopadhyay positively.

MSC 2010. Primary: 60K35, Secondary: 60J05, 82C22, 60J80.
Keywords. Recursive tree process, endogeny.
Acknowledgement. Work sponsored by grant 16-15238S of the Czech Science Foundation (GA CR).

Contents

[*]The Czech Academy of Sciences, Institute of Information Theory and Automation. Pod vodárenskou věží 4, 18208 Praha 8, Czech Republic; swart@utia.cas.cz

[†]Institute for Mathematical Stochastics, Georg-August-Universität Göttingen, Goldschmidtstr. 7, 37077 Göttingen, Germany; asturm@math.uni-goettingen.de

1 Recursive distributional equations

Let S be a Polish space and for $k \geq 1$, let S^k denote the space of all ordered sequences $(x_1, \ldots, x_k)$ of elements of S. Let $\mathcal{P}(S)$ denote the space of all probability measures on S, equipped with the topology of weak convergence and its associated Borel-σ-field. A measurable map $g : S^k \to S$ gives rise to a measurable map $\check{g} : \mathcal{P}(S)^k \to \mathcal{P}(S)$ defined as

$$\check{g}(\mu_1, \ldots, \mu_k) := \mu_1 \otimes \cdots \otimes \mu_k \circ g^{-1}, \tag{1.1}$$

where $\mu_1 \otimes \cdots \otimes \mu_k$ denotes product measure and the right-hand side of (1.1) is the image of this under the map g. A more probabilistic way to express (1.1) is to say that if $X_1, \ldots, X_k$ are independent random variables with laws $\mu_1, \ldots, \mu_k$, then $g(X_1, \ldots, X_k)$ has law $\check{g}(\mu_1, \ldots, \mu_k)$. In particular, we let $T_g : \mathcal{P}(S) \to \mathcal{P}(S)$ denote the map

$$T_g(\mu) := \check{g}(\mu, \ldots, \mu). \tag{1.2}$$

Note that T_g is in general nonlinear, unless $k = 1$.

With slight changes in the notation, the construction above works also for $k = 0$ and $k = \infty$. By definition, we let S^0 be a set containing a single element, the empty sequence, and we let S^∞ denote the space of all infinite sequences $(x_1, x_2, \ldots)$ of elements of S, equipped with the product topology and associated Borel-σ-field. It is well-known that if S is Polish, then so are S^k $(0 \leq k \leq \infty)$ and $\mathcal{P}(S)$.

Write $\mathbb{N} := \{0, 1, 2, \ldots\}$ and $\overline{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$. Let $\mathcal{G}$ be a measurable space whose elements are measurable maps $g : S^k \to S$, where $k = k_g \in \overline{\mathbb{N}}$ may depend on g, and let π be a probability law on $\mathcal{G}$. Then under suitable technical assumptions (to be made precise in the next section)

$$T(\mu) := \int_{\mathcal{G}} \pi(\mathrm{d}g)\, T_g(\mu) \qquad \big(\mu \in \mathcal{P}(S)\big) \tag{1.3}$$

defines a map $T : \mathcal{P}(S) \to \mathcal{P}(S)$. Equations of the form

$$T(\mu) = \mu \tag{1.4}$$

are called *Recursive Distributional Equations* (RDEs). A nice collection of examples of such RDEs arising in a variety of settings can be found in [AB05]. They include Galton-Watson branching processes and related random trees, probabilistic analysis of algorithms as well as statistical physics models.

2 Recursive Tree Processes

We now make our assumptions on the set $\mathcal{G}$ and probability measure π from (1.3) explicit. We fix a Polish space Ω, equipped with the Borel σ-field, modeling some source of external randomness. We let $\kappa : \Omega \to \overline{\mathbb{N}}$ be measurable, set $\Omega_k := \{\omega \in \Omega : \kappa(\omega) = k\}$ $(k \in \overline{\mathbb{N}})$, and let γ be a function such that for each $k \in \overline{\mathbb{N}}$,

$$\Omega_k \times S^k \ni (\omega, x) \mapsto \gamma[\omega](x) \in S \quad \text{is jointly measurable in } \omega \text{ and } x. \tag{2.1}$$

Note that formally, γ is a function from $\bigcup_{k \in \overline{\mathbb{N}}} \Omega_k \times S^k$ into S. We let r be a probability measure on Ω.

The joint measurability of $\gamma[\omega](x)$ in ω and x implies in particular that $\gamma[\omega] : S^{\kappa(\omega)} \to S$ is measurable for each $\omega \in \Omega$. We assume that the set $\mathcal{G}$ from (1.3) is given by

$$\mathcal{G} := \{\gamma[\omega] : \omega \in \Omega\}, \tag{2.2}$$

which we equip with the final σ-field for the map $\omega \mapsto \gamma[\omega]$. We further assume that the probability measure π from (1.3) is the image of the probability measure r on Ω under this map.

Let $\langle \mu, \phi\rangle := \int_S \mu(\mathrm{d}x)\, \phi(x)$ denote the integral of a bounded measurable function $\phi : S \to \mathbb{R}$ w.r.t. a measure μ on S. Then, under the assumptions we have just made,

$$\begin{aligned}\langle T(\mu), \phi\rangle &= \int_{\mathcal{G}} \pi(\mathrm{d}g)\, \langle T_g(\mu), \phi\rangle = \int_\Omega r(\mathrm{d}\omega)\, \langle T_{\gamma[\omega]}(\mu), \phi\rangle \\ &= \sum_{k\in\overline{\mathbb{N}}} \int_{\Omega_k} r(\mathrm{d}\omega) \int_S \mu(\mathrm{d}x_1) \cdots \int_S \mu(\mathrm{d}x_k)\, \phi\big(\gamma[\omega](x_1, \ldots, x_k)\big).\end{aligned} \tag{2.3}$$

The joint measurability of $\gamma[\omega](x)$ in $\omega \in \Omega_k$ and $x \in S^k$ guarantees that $\langle T_{\gamma[\omega]}(\mu), \phi\rangle$, which is defined by repeated integrals over S, is measurable as a function of ω and hence the integral over Ω is well-defined. Using this, one can check that our choice of the σ-field on $\mathcal{G}$ guarantees that $\langle T_g(\mu), \phi\rangle$ is measurable as a function of g, and (2.3) defines a probability measure on S.

Starting from any countable set $\mathcal{G}$ of measurable maps $g : S^k \to S$ (where $k = k_g$ may depend on g) and a probability law π on $\mathcal{G}$, it is easy to see that one can always construct Ω, r, and γ in terms of which $\mathcal{G}$ and π can then be constructed as above. For uncountable $\mathcal{G}$, the construction above not only serves as a convenient technical set-up that guarantees that the map in (1.3) is well-defined, but also has a natural interpretation, with ω playing the role of a source of external randomness. This external randomness plays a natural role in Recursive Tree Processes as introduced in [AB05], which we describe next.

For $d \in \mathbb{N}_+ := \{1, 2, \ldots\}$, let $\mathbb{T}^d$ denote the space of all finite words $\mathbf{i} = i_1 \cdots i_t$ ($t \in \mathbb{N}$) made up from the alphabet $\{1, \ldots, d\}$, and define $\mathbb{T}^\infty$ similarly, using the alphabet $\mathbb{N}_+$. Let $\varnothing$ denote the word of length zero. We view $\mathbb{T}^d$ as a tree with root $\varnothing$, where each vertex $\mathbf{i} \in \mathbb{T}^d$ has d children $\mathbf{i}1, \mathbf{i}2, \ldots$, and each vertex $\mathbf{i} = i_1 \cdots i_t$ except the root has precisely one ancestor $i_1 \cdots i_{t-1}$. If $\mathbf{i}, \mathbf{j} \in \mathbb{T}^d$ with $\mathbf{i} = i_1 \cdots i_s$ and $\mathbf{j} = j_1 \cdots j_t$, then we define the concatenation $\mathbf{ij} \in \mathbb{T}^d$ by $\mathbf{ij} = i_1 \cdots i_s j_1 \cdots j_t$. We denote the length of a word $\mathbf{i} = i_1 \cdots i_t$ by $|\mathbf{i}| := t$ and set $\mathbb{T}^d_{(t)} := \{\mathbf{i} \in \mathbb{T}^d : |\mathbf{i}| < t\}$. For any subtree $\mathbb{U} \subset \mathbb{T}$, we let $\partial\mathbb{U} := \{\mathbf{i} = i_1 \cdots i_{|\mathbf{i}|} \in \mathbb{T}^d : i_1 \cdots i_{|\mathbf{i}|-1} \in \mathbb{U},\ \mathbf{i} \notin \mathbb{U}\}$ denote the outer boundary of $\mathbb{U}$. In particular, $\partial\mathbb{T}^d_{(t)} = \{\mathbf{i} \in \mathbb{T}^d : |\mathbf{i}| = t\}$ is the set of all vertices at distance t from the root.

Fix some $d \in \overline{\mathbb{N}}_+ := \mathbb{N}_+ \cup \{\infty\}$ such that $\kappa(\omega) \le d$ for all $\omega \in \Omega$, and to simplify notation write $\mathbb{T} := \mathbb{T}^d$. Let $(\omega_\mathbf{i})_{\mathbf{i}\in\mathbb{T}}$ be an i.i.d. collection of Ω-valued random variables with common law r. Fix $t \ge 1$ and $\mu \in \mathcal{P}(S)$, and let $(X_\mathbf{i})_{\mathbf{i}\in\partial\mathbb{T}_{(t)}}$ be a collection of S-valued random variables such that

$$(X_\mathbf{i})_{\mathbf{i}\in\partial\mathbb{T}_{(t)}} \text{ are i.i.d. with common law } \mu \text{ and independent of } (\omega_\mathbf{i})_{\mathbf{i}\in\mathbb{T}_{(t)}}. \tag{2.4}$$

Define $(X_\mathbf{i})_{\mathbf{i}\in\mathbb{T}_{(t)}}$ inductively by[1]

$$X_\mathbf{i} := \gamma[\omega_\mathbf{i}](X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}\kappa(\omega_\mathbf{i})}) \tag{2.5}$$

[1] Here and in similar formulas to come, it is understood that the notation should be suitably adapted if $\kappa(\omega_\mathbf{i}) = \infty$, e.g., in this place, $X_\mathbf{i} = \gamma[\omega_\mathbf{i}]\big(X_{\mathbf{i}1}, X_{\mathbf{i}2}, \ldots\big)$.

Then it is easy to see that for each $1 \leq s \leq t$,

$$(X_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(s)}} \text{ are i.i.d. with common law } T^{t-s}(\mu) \text{ and independent of } (\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_{(s)}}, \tag{2.6}$$

where T^n denotes the n-th iterate of the map in (1.3). Also, $X_\varnothing$ (the state at the root) has law $T^t(\mu)$. If μ is a solution of the RDE (1.4), then, by Kolmogorov's extension theorem there exists a collection $(\omega_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ of random variables whose joint law is uniquely characterized by the following requirements:

$$\begin{aligned} &\text{(i)} \quad (\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}} \text{ is an i.i.d. collection of } \Omega\text{-valued r.v.'s with common law } r, \\ &\text{(ii)} \quad \text{for each } t \geq 1, \text{ the } (X_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(t)}} \text{ are i.i.d. with common law } \mu \\ &\qquad \text{and independent of } (\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_{(t)}}, \\ &\text{(iii)} \quad X_{\mathbf{i}} := \gamma[\omega_{\mathbf{i}}](X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}\kappa(\omega_{\mathbf{i}})}) \qquad (\mathbf{i} \in \mathbb{T}). \end{aligned} \tag{2.7}$$

We call such a collection $(\omega_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ a *Recursive Tree Process* (RTP) corresponding to the map γ and the solution μ of the RDE (1.4). We can think of an RTP as a generalization of a stationary and time reversed Markov chain, where the time index set $\mathbb{T}$ has a tree structure and time flows in the direction of the root. In each step, the new value $X_{\mathbf{i}}$ is a function of the previous values $X_{\mathbf{i}1}, X_{\mathbf{i}2} \ldots$, plus some independent randomness, represented by the random variables $(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$. Following [AB05, Def 7], we say that the RTP corresponding to a solution μ of the RDE (1.4) is *endogenous* if $X_\varnothing$ is measurable w.r.t. the σ-field generated by the random variables $(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$.

Endogeny is somewhat similar to pathwise uniqueness of stochastic differential equations, in the sense that it asks whether given $(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$, there always exists a "strong solution" $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ on the same probability space, or whether on the other hand additional randomness is needed to construct $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$. Since for each $t \geq 1$, $X_\varnothing$ is a function of $(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_{(t)}}$ and the "boundary conditions" $(X_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(t)}}$, endogeny says that in a certain almost sure sense, the effect of the boundary conditions disappears as $t \to \infty$. Nevertheless, endogeny does not imply uniqueness of solutions to the RDE (1.4). Indeed, it is possible for a RDE to have several solutions, while some of the corresponding RTPs are endogenous and others are not. In the special case that $\mathbb{T} = \mathbb{T}^1$, an RTP is a time reversed stationary Markov chain $\ldots, X_{11}, X_1, X_\varnothing$ generated by i.i.d. random variables $\ldots, \omega_{11}, \omega_1$. In this context, equivalent formulations of endogeny have been investigated in the literature, see for example [BL07] who point back to [Ros59]. Endogeny also plays a role, for example, in the coupling from the past algorithm by Propp and Wilson [PW96]. We point to Lemma 14 and 15 of Section 2.6 in [AB05] for an analogous statement on tree-structured coupling from the past.

Endogeny of RTPs is related to, but not the same as triviality of the tail-σ-field as defined for infinite volume Gibbs measures. Indeed, if $(\omega_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ is an RTP, then the law of $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ is a Gibbs measure on $S^{\mathbb{T}}$. The *tail-σ-field* of such a Gibbs measure is defined as

$$\mathcal{T} := \bigcap_{t\geq 1} \sigma\big((X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}\setminus\mathbb{T}_{(t)}}\big). \tag{2.8}$$

It is known that if $(\omega_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ is endogenous, then the tail-σ-field of $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ is trivial [Ant06, Prop. 1], but the converse implication does not hold [Ant06, Example 1]. It is known that triviality of the tail-σ-field is equivalent to nonreconstructability in information theory [Mos01, Prop. 15], and also to extremality of $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ as a Gibbs measure [Geo11, Section 7.1].

3 The n-variate RDE

Let $g : S^k \to S$ with $k \geq 0$ be a measurable map and let $n \geq 1$ be an integer. We can naturally identify the space $(S^n)^k$ with the space of all $n \times k$ matrices $x = (x_i^j)_{i=1,\ldots,k}^{j=1,\ldots,n}$. We let $x^j := (x_1^j, \ldots, x_k^j)$ and $x_i = (x_i^1, \ldots, x_i^n)$ denote the rows and columns of such a matrix, respectively. With this notation, we define an *n-variate map* $g^{(n)} : (S^n)^k \to S^n$ by

$$g^{(n)}(x) := \big(g(x^1), \ldots, g(x^n)\big) \qquad \big(x \in (S^n)^k\big). \tag{3.1}$$

This notation is easily generalized to $k = \infty$ or $n = \infty$, or both. The map $g^{(n)}$ describes n systems that are coupled in such a way that the same map g is applied to each system. We will be interested in the *n-variate map* (compare (1.3))

$$T^{(n)}(\nu) := \int_{\mathcal{G}} \pi(\mathrm{d}g)\, T_{g^{(n)}}(\nu) \qquad \big(\nu \in \mathcal{P}(S^n)\big). \tag{3.2}$$

and the corresponding *n-variate RDE* (compare (1.4))

$$T^{(n)}(\nu) = \nu. \tag{3.3}$$

If γ satisfies (2.1), then the same is true for $\gamma^{(n)}$ so $T^{(n)}(\nu)$ is well-defined. The maps $T^{(n)}$ are consistent in the following sense. Let $\nu|_{\{i_1,\ldots,i_m\}}$ denote the marginal of ν with respect to the coordinates $i_1, \ldots, i_m$, i.e., the image of ν under the projection $(x_1, \ldots, x_n) \mapsto (x_{i_1}, \ldots, x_{i_m})$. Then

$$T^{(n)}(\nu)\big|_{\{i_1,\ldots,i_m\}} = T^{(m)}\big(\nu\big|_{\{i_1,\ldots,i_m\}}\big). \tag{3.4}$$

In particular, if ν solves the n-variate RDE (3.3), then its one-dimensional marginals $\nu|_{\{m\}}$ $(1 \leq m \leq n)$ solve the RDE (1.4). For any $\mu \in \mathcal{P}(S)$, we let

$$\mathcal{P}(S^n)_\mu := \big\{\nu \in \mathcal{P}(S^n) : \nu|_{\{m\}} = \mu \ \forall 1 \leq m \leq n\big\} \tag{3.5}$$

denote the set of probability measures on S^n whose one-dimensional marginals are all equal to μ. We also let $\mathcal{P}_{\rm sym}(S^n)$ denote the space of all probability measures on S^n that are symmetric with respect to permutations of the coordinates $\{1, \ldots, n\}$, and denote $\mathcal{P}_{\rm sym}(S^n)_\mu := \mathcal{P}_{\rm sym}(S^n) \cap \mathcal{P}(S^n)_\mu$. It is easy to see that $T^{(n)}$ maps $\mathcal{P}_{\rm sym}(S^n)$ into itself. If μ solves the RDE (1.4), then $T^{(n)}$ also maps $\mathcal{P}_{\rm sym}(S^n)_\mu$ into itself.

Given a measure $\mu \in \mathcal{P}(S)$, we define $\overline{\mu}^{(n)} \in \mathcal{P}(S^n)$ by

$$\overline{\mu}^{(n)} := \mathbb{P}\big[(X, \ldots, X) \in \cdot\,\big] \quad \text{where } X \text{ has law } \mu. \tag{3.6}$$

We will prove the following theorem, which is similar to [AB05, Thm 11]. The main improvement compared to the latter is that the implication (ii)⇒(i) is shown without the additional assumption that $T^{(2)}$ is continuous with respect to weak convergence, solving Open Problem 12 of [AB05]. We have learned that this problem has been solved before using an argument from [BL07], although its solution has not been published. We refer to Appendix B for a comparison of our solution and this other solution. Below, ⇒ denotes weak convergence of probability measures.

Theorem 1 (Endogeny and bivariate uniqueness) *Let μ be a solution to the RDE (1.4). Then the following statements are equivalent.*

- (i) *The RTP corresponding to μ is endogenous.*
- (ii) *The measure $\overline{\mu}^{(2)}$ is the unique fixed point of $T^{(2)}$ in the space $\mathcal{P}_{\rm sym}(S^2)_\mu$.*
- (iii) *$(T^{(n)})^t(\nu) \underset{t\to\infty}{\Longrightarrow} \overline{\mu}^{(n)}$ for all $\nu \in \mathcal{P}(S^n)_\mu$ and $n \in \overline{\mathbb{N}}_+$.*

4 The higher-level RDE

In this section we introduce a higher-level map $\check{T}$ that through its n-th moment measures contains all n-variate maps (Lemma 2 below). In the next section, we will use this higher-level map to give a short and elegant proof of Theorem 1. We believe the methods of the present section to be of wider interest. In particular, in future work we plan to use them to study iterates of the n-variate maps for a non-endogenous RTP related to systems with cooperative branching.

Let ξ be a random probability measure on S, i.e., a $\mathcal{P}(S)$-valued random variable, and let $\rho \in \mathcal{P}(\mathcal{P}(S))$ denote the law of ξ. Conditional on ξ, let $X^1, \ldots, X^n$ be independent with law ξ. Then (see Lemma 7 below)

$$\rho^{(n)} := \mathbb{P}\big[(X^1, \ldots, X^n) \in \cdot\,\big] = \mathbb{E}\big[\underbrace{\xi \otimes \cdots \otimes \xi}_{n \text{ times}}\big] \tag{4.1}$$

is called the *n-th moment measure* of ξ. Here, the expectation of a random measure ξ on S is defined in the usual way, i.e., $\mathbb{E}[\xi]$ is the deterministic measure defined by $\int \phi \,\mathrm{d}\mathbb{E}[\xi] := \mathbb{E}[\int \phi \,\mathrm{d}\xi]$ for any bounded measurable $\phi : S \to \mathbb{R}$. A similar definition applies to measures on S^n. With slight changes in the notation, $\rho^{(n)}$ can also be defined for $n = \infty$.

We observe that $\rho^{(n)} \in \mathcal{P}_{\rm sym}(S^n)$ for each $\rho \in \mathcal{P}(\mathcal{P}(S))$ and $n \in \overline{\mathbb{N}}_+$. By De Finetti's theorem, for $n = \infty$ the converse implication also holds. Indeed, $\mathcal{P}_{\rm sym}(S^\infty)$ is the space of exchangeable probability measures on S^∞ and De Finetti says that each element of $\mathcal{P}_{\rm sym}(S^\infty)$ is of the form $\rho^{(\infty)}$ for some $\rho \in \mathcal{P}(\mathcal{P}(S))$. Thus, we have a natural identification $\mathcal{P}_{\rm sym}(S^\infty) \cong \mathcal{P}(\mathcal{P}(S))$ and through this identification the map $T^{(\infty)} : \mathcal{P}_{\rm sym}(S^\infty) \to \mathcal{P}_{\rm sym}(S^\infty)$ corresponds to a map on $\mathcal{P}(\mathcal{P}(S))$. Our next aim is to identify this map.

Let $\check{T} : \mathcal{P}(\mathcal{P}(S)) \to \mathcal{P}(\mathcal{P}(S))$ be given by

$$\check{T}(\rho) := \int_{\mathcal{G}} \pi(\mathrm{d}g)\, T_{\check{g}}(\rho) \qquad \big(\rho \in \mathcal{P}(\mathcal{P}(S))\big), \tag{4.2}$$

where for any $g : S^k \to S$, the map $\check{g} : \mathcal{P}(S)^k \to \mathcal{P}(S)$ is defined as in (1.1). If γ satisfies (2.1), then the same is true for $\check{\gamma}$ so $\check{T}(\rho)$ is well-defined. Note that $T_{\check{g}}(\rho) = \check{\check{g}}(\rho, \ldots, \rho)$ by (1.2). We call $\check{T}$ the *higher-level map*, which gives rise to the *higher-level RDE*

$$\check{T}(\rho) = \rho. \tag{4.3}$$

The following lemma shows that $\check{T}$ is the map corresponding to $T^{(\infty)}$ we were looking for. More generally, the lemma links $\check{T}$ to the n-variate maps $T^{(n)}$.

Lemma 2 (Moment measures) *Let $n \in \overline{\mathbb{N}}_+$ and let $T^{(n)}$ and $\check{T}$ be defined as in (3.2) and (4.2). Then the n-th moment measure of $\check{T}(\rho)$ is given by*

$$\check{T}(\rho)^{(n)} = T^{(n)}(\rho^{(n)}) \qquad \big(\rho \in \mathcal{P}(\mathcal{P}(S))\big). \tag{4.4}$$

Lemma 2 implies in particular that if ρ solves the higher-level RDE (4.3), then its first moment measure $\rho^{(1)}$ solves the original RDE (1.4). For any $\mu \in \mathcal{P}(S)$, we let

$$\mathcal{P}(\mathcal{P}(S))_\mu := \big\{\rho \in \mathcal{P}(\mathcal{P}(S)) : \rho^{(1)} = \mu\big\} \tag{4.5}$$

denote the set of all ρ whose first moment measure is μ. Note that $\rho \in \mathcal{P}(\mathcal{P}(S))_\mu$ implies $\rho^{(n)} \in \mathcal{P}_{\rm sym}(S^n)_\mu$ for each $n \geq 1$.

We equip $\mathcal{P}(\mathcal{P}(S))$ with the *convex order*. By Theorem 13 in Appendix A, two measures $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(S))$ are ordered in the convex order, denoted $\rho_1 \leq_{\rm cv} \rho_2$, if and only if there exists an S-valued random variable X defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and sub-σ-fields $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}$ such that $\rho_i = \mathbb{P}\big[\mathbb{P}[X \in \cdot\,|\mathcal{F}_i] \in \cdot\,\big]$ $(i = 1, 2)$. It is not hard to see that $\mathcal{P}(\mathcal{P}(S))_\mu$ has a minimal and maximal element w.r.t. the convex order. For any $\mu \in \mathcal{P}(S)$, let us define

$$\overline{\mu} := \mathbb{P}\big[\delta_X \in \cdot\,\big] \quad \text{where } X \text{ has law } \mu. \tag{4.6}$$

Clearly $\delta_\mu, \overline{\mu} \in \mathcal{P}(\mathcal{P}(S))_\mu$. Moreover (as will be proved in Section 6 below)

$$\delta_\mu \leq_{\rm cv} \rho \leq_{\rm cv} \overline{\mu} \quad \text{for all} \quad \rho \in \mathcal{P}(\mathcal{P}(S))_\mu. \tag{4.7}$$

In line with notation that has already been introduced in (3.6), the n-th moment measures of δ_μ and $\overline{\mu}$ are given by

$$\delta_\mu^{(n)} = \mathbb{P}\big[(X_1, \ldots, X_n) \in \cdot\,\big] \quad \text{and} \quad \overline{\mu}^{(n)} = \mathbb{P}\big[(X, \ldots, X) \in \cdot\,\big], \tag{4.8}$$

where $X_1, \ldots, X_n$ are i.i.d. with common law μ and X has law μ. The following proposition says that the higher-level RDE (4.3) has a minimal and maximal solution with respect to the convex order.

Proposition 3 (Minimal and maximal solutions) *The map $\check{T}$ is monotone w.r.t. the convex order. Let μ be a solution to the RDE (1.4). Then $\check{T}$ maps $\mathcal{P}(\mathcal{P}(S))_\mu$ into itself. There exists a unique $\underline{\mu} \in \mathcal{P}(\mathcal{P}(S))_\mu$ such that*

$$\check{T}^t(\delta_\mu) \underset{t\to\infty}{\Longrightarrow} \underline{\mu}, \tag{4.9}$$

where $\Rightarrow$ denotes weak convergence of measures on $\mathcal{P}(S)$, equipped with the topology of weak convergence. The measures $\underline{\mu}$ and $\overline{\mu}$ solve the higher-level RDE (4.3), and any $\rho \in \mathcal{P}(\mathcal{P}(S))_\mu$ that solves the higher-level RDE (4.3) must satisfy

$$\underline{\mu} \leq_{\rm cv} \rho \leq_{\rm cv} \overline{\mu}. \tag{4.10}$$

Since $\underline{\mu}$ and $\overline{\mu}$ solve the higher-level RDE (4.3), there exist RTPs corresponding to $\underline{\mu}$ and $\overline{\mu}$. The following proposition gives an explicit description of these higher-level RTPs.

Proposition 4 (Higher-level RTPs) *Let $(\omega_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be an RTP corresponding to a solution μ of the RDE (1.4). Set*

$$\xi_{\mathbf{i}} := \mathbb{P}\big[X_{\mathbf{i}} \in \cdot\,|\,(\omega_{\mathbf{ij}})_{\mathbf{j}\in\mathbb{T}}\big]. \tag{4.11}$$

Then $(\omega_{\mathbf{i}}, \xi_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ is an RTP corresponding to the map $\check{\gamma}$ and the solution $\underline{\mu}$ of the higher-level RDE (4.3). Also, $(\omega_{\mathbf{i}}, \delta_{X_{\mathbf{i}}})_{\mathbf{i}\in\mathbb{T}}$ is an RTP corresponding to the map $\check{\gamma}$ and $\overline{\mu}$.

In general, we can interpret the higher-level map $\check{T}$ as follows. Fix $t \geq 1$, and let $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(t)}}$ be i.i.d. random variables, independent of $(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_{(t)}}$, where the $X_{\mathbf{i}}$'s take values in S and the $Y_{\mathbf{i}}$'s take values in some arbitrary measurable space. Define $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_{(t)}}$ inductively as in (2.5) and for $\mathbf{i} \in \mathbb{T}_{(t)} \cup \partial\mathbb{T}_{(t)}$, let $\xi_{\mathbf{i}}$ denote the conditional law of $X_{\mathbf{i}}$ given $(\omega_{\mathbf{ij}})_{\mathbf{ij}\in\mathbb{T}_{(t)}}$ and $(Y_{\mathbf{ij}})_{\mathbf{ij}\in\partial\mathbb{T}_{(t)}}$. Then the $(\xi_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(t)}}$ are i.i.d. with some common law $\rho \in \mathcal{P}(\mathcal{P}(S))$. Using Lemma 8 below, it is not hard to see that for each $1 \leq s \leq t$,

$$(\xi_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(s)}} \text{ are i.i.d. with common law } \check{T}^{t-s}(\rho) \text{ and independent of } (\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_{(s)}}. \tag{4.12}$$

Let μ denote the common law of the random variables $(X_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(t)}}$. We think of the random variables $(Y_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(t)}}$ as providing extra information about the $(X_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(t)}}$. The convex order measures how much extra information we have. For $\xi_{\mathbf{i}}=\delta_{X_{\mathbf{i}}}$, we have perfect information, while on the other hand for $\xi_{\mathbf{i}}=\mu$ the $Y_{\mathbf{i}}$'s provided no extra information. A solution to the higher-level RDE gives rise to a higher-level RTP that can be interpreted as a normal (low-level) RTP $(\omega_{\mathbf{i}},X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ in which we have extra information about the $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$. The solutions $\underline{\mu}$ and $\overline{\mu}$ of the higher-level RDE correspond to minimal and maximal knowledge about $X_{\mathbf{i}}$, respectively, where either we know only $(\omega_{\mathbf{ij}})_{\mathbf{j}\in\mathbb{T}}$, or we have full knowledge about $X_{\mathbf{i}}$.

We will derive Theorem 1 from the following theorem, which is our main result.

Theorem 5 (The higher-level RDE) *Let μ be a solution to the RDE (1.4). Then the following statements are equivalent.*

(i) *The RTP corresponding to μ is endogenous.*

(ii) $\underline{\mu}=\overline{\mu}$.

(iii) $\check{T}^t(\rho)\underset{t\to\infty}{\Longrightarrow}\overline{\mu}$ *for all $\rho\in\mathcal{P}(\mathcal{P}(S))_\mu$.*

5 Proof of the main theorem

In this section, we use Lemma 2 and Propositions 3 and 4 to prove Theorems 1 and 5. We need one more lemma.

Lemma 6 (Convergence in probability) *Let $(\omega_{\mathbf{i}},X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be an endogenous RTP corresponding to a solution μ of the RDE (1.4), and let $(Y_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be an independent i.i.d. collection of S-valued random variables with common law μ. For each $t\geq 1$, set $X^t_{\mathbf{i}}:=Y_{\mathbf{i}}$ $(\mathbf{i}\in\partial\mathbb{T}_{(t)})$, and define $(X^t_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_{(t)}}$ inductively by*

$$X^t_{\mathbf{i}}=\gamma[\omega_{\mathbf{i}}](X^t_{\mathbf{i}1},\ldots,X^t_{\mathbf{i}\kappa(\omega_{\mathbf{i}})})\qquad(\mathbf{i}\in\mathbb{T}_{(t)}).\tag{5.1}$$

Then

$$X^t_\varnothing\underset{t\to\infty}{\longrightarrow}X_\varnothing\quad\text{in probability.}\tag{5.2}$$

Proof The argument is basically the same as in the proof of [AB05, Thm 11 (c)], but for completeness, we give it here. Let $f,g:S\to\mathbb{R}$ be bounded and continuous and let $\mathcal{F}_t$ resp. $\mathcal{F}_\infty$ be the σ-fields generated by $(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_{(t)}}$ resp. $(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$. Since $X_\varnothing$ and $X^t_\varnothing$ are conditionally independent and identically distributed given $\mathcal{F}_t$,

$$\begin{aligned}&\mathbb{E}\big[f(X_\varnothing)g(X^t_\varnothing)\big]=\mathbb{E}\big[\mathbb{E}\big[f(X_\varnothing)\,\big|\,\mathcal{F}_t\big]\mathbb{E}\big[g(X^t_\varnothing)\,\big|\,\mathcal{F}_t\big]\big]\\&\quad=\mathbb{E}\big[\mathbb{E}\big[f(X_\varnothing)\,\big|\,\mathcal{F}_t\big]\mathbb{E}\big[g(X_\varnothing)\,\big|\,\mathcal{F}_t\big]\big]\\&\quad\underset{t\to\infty}{\longrightarrow}\mathbb{E}\big[\mathbb{E}\big[f(X_\varnothing)\,\big|\,\mathcal{F}_\infty\big]\mathbb{E}\big[g(X_\varnothing)\,\big|\,\mathcal{F}_\infty\big]\big]=\mathbb{E}\big[f(X_\varnothing)g(X_\varnothing)\big],\end{aligned}\tag{5.3}$$

where we have used martingale convergence and in the last step also endogeny. Since this holds for arbitrary f,g, we conclude that the law of $(X_\varnothing,X^t_\varnothing)$ converges weakly to the law of $(X_\varnothing,X_\varnothing)$, which implies (5.2). ∎ ∎

Proof of Theorem 5 If the RTP corresponding to μ is endogenous, then the random variable $\xi_\varnothing$ defined in (4.11) satisfies $\xi_\varnothing=\delta_{X_\varnothing}$. By Proposition 4, $\xi_\varnothing$ and $\delta_{X_\varnothing}$ have laws $\underline{\mu}$ and $\overline{\mu}$, respectively, so (i)$\Rightarrow$(ii). Conversely, if (i) does not hold, then $\xi_\varnothing$ is with positive probability not a delta measure, so (i)$\Leftrightarrow$(ii).

The implication (iii)$\Rightarrow$(ii) is immediate from the definition of $\underline{\mu}$ in (4.9). To get the converse implication, we observe that by Proposition 3, $\check{T}$ is monotone with respect to the convex order, so (4.7) implies

$$\check{T}^t(\delta_\mu) \leq_{\rm cv} \check{T}^t(\rho) \leq_{\rm cv} \check{T}^t(\overline{\mu}) \qquad (t \geq 0). \tag{5.4}$$

By Proposition 3, $\check{T}$ maps $\mathcal{P}(\mathcal{P}(S))_\mu$ into itself, so $\check{T}^t(\rho) \in \mathcal{P}(\mathcal{P}(S))_\mu$ for each $t \geq 0$, and hence by Lemma 9 in Appendix A, the measures $(\check{T}^t(\rho))_{t\geq 1}$ are tight. By Proposition 3, the left-hand side of (5.4) converges weakly to $\underline{\mu}$ as $t \to \infty$ while the right-hand side equals $\overline{\mu}$ for each t, so we obtain that any subsequential limit $\check{T}^{t_n}(\rho) \Rightarrow \rho_*$ satisfies $\underline{\mu} \leq_{\rm cv} \rho_* \leq_{\rm cv} \overline{\mu}$. In particular, this shows that (ii)$\Rightarrow$(iii). ∎ ∎

Proof of Theorem 1 The implication (iii)$\Rightarrow$(ii) is trivial. By Lemma 2 and the fact that $\underline{\mu}$ and $\overline{\mu}$ solve the higher-level RDE, we see that (ii) implies $\underline{\mu}^{(2)} = \overline{\mu}^{(2)}$. By Proposition 3, $\underline{\mu} \leq_{\rm cv} \overline{\mu}$. Now Lemma 14 from Appendix A shows that $\underline{\mu}^{(2)} = \overline{\mu}^{(2)}$ and $\underline{\mu} \leq_{\rm cv} \overline{\mu}$ imply $\underline{\mu} = \overline{\mu}$, so applying Theorem 5 we obtain that (ii)$\Rightarrow$(i).

To complete the proof, we will show that (i)$\Rightarrow$(iii). Let $(\omega_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be an RTP corresponding to μ and let $(Y^1_{\mathbf{i}}, \ldots, Y^n_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be an independent i.i.d. collection of S^n-valued random variables with common law ν. For each $t \geq 1$ and $1 \leq m \leq n$, set $X^{m,t}_{\mathbf{i}} := Y^m_{\mathbf{i}} \quad (\mathbf{i} \in \partial\mathbb{T}_{(t)})$, and define $(X^{m,t}_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_{(t)}}$ inductively as in (5.1). Then $(X^{1,t}_\varnothing, \ldots X^{n,t}_\varnothing)$ has law $(T^{(n)})^t(\nu)$, and using endogeny, Lemma 6 tells us that

$$(X^{1,t}_\varnothing, \ldots X^{n,t}_\varnothing) \underset{t\to\infty}{\longrightarrow} (X_\varnothing, \ldots, X_\varnothing) \quad \text{in probability.} \tag{5.5}$$

Since the right-hand side has law $\overline{\mu}^{(n)}$ (recall (4.8)), this completes the proof. With a slight change of notation, this argument also works for $n = \infty$. ∎ ∎

6 Other proofs

In this section, we provide the proofs of Lemma 2 and Propositions 3 and 4, as well as formula (4.7). We start with some preliminary observations.

Lemma 7 (Moment measures) *Let $X^1, \ldots, X^n$ be S-valued random variables such that conditionally on some σ-field $\mathcal{H}$, the $X^1, \ldots, X^n$ are i.i.d. with (random) law $\mathbb{P}[X^j \in \cdot\,|\mathcal{H}] = \xi$ a.s. $(j = 1, \ldots, n)$. Let $\rho \in \mathcal{P}(\mathcal{P}(S))$ denote the law of ξ, i.e.,*

$$\rho = \mathbb{P}\big[\mathbb{P}[X^j \in \cdot\,|\mathcal{H}] \in \cdot\,\big] \qquad (j = 1, \ldots, n). \tag{6.1}$$

Then

$$\rho^{(n)} = \mathbb{P}\big[(X^1, \ldots, X^n) \in \cdot\,\big] = \mathbb{E}\big[\underbrace{\xi \otimes \cdots \otimes \xi}_{n \text{ times}}\big]. \tag{6.2}$$

Proof This follows by writing

$$\begin{aligned}\mathbb{E}\big[\prod_{i=1}^n f_i(X^i)\big] &= \mathbb{E}\big[\prod_{i=1}^n \mathbb{E}[f_i(X^i)\,|\,\mathcal{H}]\big] \\ &= \int \rho(\mathrm{d}\xi) \int_{S^n} \xi(\mathrm{d}x_1)\cdots\xi(\mathrm{d}x_n)\, f_1(x)\cdots f_n(x_n).\end{aligned} \tag{6.3}$$

for arbitrary bounded measurable $f_i : S \to \mathbb{R}$. ∎ ∎

Lemma 8 (Higher-level map) *Let $X_1, \ldots, X_k$ and $\mathcal{H}_1, \ldots, \mathcal{H}_k$ be S-valued random variables and σ-fields, respectively, such that $(X_1, \mathcal{H}_1), \ldots, (X_k, \mathcal{H}_k)$ are i.i.d. Let*

$$\rho = \mathbb{P}\big[\mathbb{P}[X_i \in \cdot\,|\mathcal{H}_i] \in \cdot\,\big] \qquad (i = 1, \ldots, k). \tag{6.4}$$

Let $\mathcal{H}_1 \vee \cdots \vee \mathcal{H}_k$ denote the σ-field generated by $\mathcal{H}_1, \ldots, \mathcal{H}_k$. Then, for each measurable $g : S^k \to S$,

$$T_{\check{g}}(\rho) = \mathbb{P}\big[\mathbb{P}[g(X_1, \ldots, X_k) \in \cdot\,|\mathcal{H}_1 \vee \cdots \vee \mathcal{H}_k] \in \cdot\,\big]. \tag{6.5}$$

Similarly, if ω is an independent Ω-valued random variable with law r and $\mathcal{F}$ is the σ-algebra generated by ω, then

$$\check{T}(\rho) = \mathbb{P}\big[\mathbb{P}[\gamma[\omega](X_1, \ldots, X_k) \in \cdot\,|\mathcal{H}_1 \vee \cdots \vee \mathcal{H}_k \vee \mathcal{F}] \in \cdot\,\big]. \tag{6.6}$$

Proof Let $\xi_i := \mathbb{P}[X_i \in \cdot\,|\mathcal{H}_i]$. Then (i) $\xi_1, \ldots, \xi_k$ are i.i.d. with common law ρ, and (ii) conditional on $\mathcal{H}_1 \vee \cdots \vee \mathcal{H}_k$, the random variables $X_1, \ldots, X_k$ are independent with laws $\xi_1, \ldots, \xi_k$, respectively. Now (ii) implies

$$\mathbb{P}[g(X_1, \ldots, X_k) \in \cdot\,|\mathcal{H}_1 \vee \cdots \vee \mathcal{H}_k] = \check{g}(\xi_1, \ldots, \xi_k) \quad \text{a.s.} \tag{6.7}$$

and in view of (i), (6.5) follows. Since ω is independent of $(X_1, \mathcal{H}_1), \ldots, (X_k, \mathcal{H}_k)$, conditional on $\mathcal{H}_1 \vee \cdots \vee \mathcal{H}_k \vee \mathcal{F}$, the random variables $X_1, \ldots, X_k$ are again independent with laws $\xi_1, \ldots, \xi_k$, respectively, and hence

$$\mathbb{P}[\gamma[\omega](X_1, \ldots, X_k) \in \cdot\,|\mathcal{H}_1 \vee \cdots \vee \mathcal{H}_k \vee \mathcal{F}] = \check{\gamma}[\omega](\xi_1, \ldots, \xi_k) \quad \text{a.s.}, \tag{6.8}$$

which implies (6.6). With a slight change in notation, these formulas hold also for $k = \infty$. ∎

Proof of formula (4.7) Let ξ be a $\mathcal{P}(S)$-valued random variable with law ρ and conditional on ξ, let X be an S-valued random variable with law ξ. Let $\mathcal{F}_0$ be the trivial σ-field, let $\mathcal{F}_1$ be the σ-field generated by ξ, and let $\mathcal{F}_2$ be the σ-field generated by ξ and X. Then $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2$. Since $\rho^{(1)} = \mu$, the random variable X has law μ. Now

$$\begin{aligned}
\mathbb{P}\big[\mathbb{P}[X \in \cdot\,|\,\mathcal{F}_0] \in \cdot\,\big] &= \mathbb{P}\big[\mu \in \cdot\,\big] = \delta_\mu, \\
\mathbb{P}\big[\mathbb{P}[X \in \cdot\,|\,\mathcal{F}_1] \in \cdot\,\big] &= \mathbb{P}\big[\xi \in \cdot\,\big] = \rho, \\
\mathbb{P}\big[\mathbb{P}[X \in \cdot\,|\,\mathcal{F}_2] \in \cdot\,\big] &= \mathbb{P}\big[\delta_X \in \cdot\,\big] = \overline{\mu}.
\end{aligned} \tag{6.9}$$

This proves that $\delta_\mu \leq_{\rm cv} \rho \leq_{\rm cv} \overline{\mu}$. ∎ ∎

Proof of Lemma 2 Let $\xi_1, \ldots, \xi_k$ be i.i.d. with common law ρ and conditional on $\xi_1, \ldots, \xi_k$, let $(X_i^j)_{i=1,\ldots,k}^{j=1,\ldots,n}$ be independent S-valued random variables such that X_i^j has law ξ_i. Let $\mathcal{H}_i$ denote the σ-field generated by ξ_i. Then $\rho = \mathbb{P}[\mathbb{P}[X_i^j \in \cdot\,|\mathcal{H}_i] \in \cdot\,]$ for each i, j. By (6.2),

$$\rho^{(n)} = \mathbb{P}\big[(X_i^1, \ldots, X_i^n) \in \cdot\,\big] \qquad (i = 1, \ldots, k). \tag{6.10}$$

Set $X_i := (X_i^1, \ldots, X_i^n)$ and $X^j := (X_1^j, \ldots, X_k^j)$. Since $X_1, \ldots, X_k$ are independent with law $\rho^{(n)}$,

$$T_{g^{(n)}}(\rho^{(n)}) = \mathbb{P}\big[g^{(n)}(X_1, \ldots, X_k) \in \cdot\,\big] = \mathbb{P}\big[\big(g(X^1), \ldots, g(X^n)\big) \in \cdot\,\big]. \tag{6.11}$$

Let $\mathcal{H} := \mathcal{H}_1 \vee \cdots \vee \mathcal{H}_k$. Since $(X_1^j, \mathcal{H}_1), \ldots, (X_k^j, \mathcal{H}_k)$ are i.i.d. for each j, formula (6.5) tells us that $T_{\check{g}}(\rho) = \mathbb{P}\big[\mathbb{P}[g(X^j) \in \cdot\,|\mathcal{H}] \in \cdot\,\big]$ $(j = 1, \ldots, n)$. Since conditionally on $\mathcal{H}$, the $g(X^1), \ldots, g(X^n)$ are i.i.d., formula (6.2) tells us that

$$T_{\check{g}}(\rho)^{(n)} = \mathbb{P}\big[\big(g(X^1), \ldots, g(X^n)\big) \in \cdot\,\big]. \tag{6.12}$$

Combining this with (6.11), we see that $T_{\check{g}}(\rho)^{(n)} = T_{g^{(n)}}(\rho^{(n)})$ for each $g \in \mathcal{G}$. Now (4.4) follows by integrating w.r.t. π. ∎ ∎

Proof of Propositions 3 and 4 We first show that $\check{T}$ is monotone w.r.t. the convex order. Let $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(S))_\mu$ satisfy $\rho_1 \leq_{\rm cv} \rho_2$. Then we can construct S-valued random variables $X_1, \ldots, X_k$ as well as σ-fields $(\mathcal{H}_i^j)_{i=1,\ldots,k}^{j=1,2}$, such that $(X_1, \mathcal{H}_1^1, \mathcal{H}_1^2), \ldots, (X_k, \mathcal{H}_k^1, \mathcal{H}_k^2)$ are i.i.d.,

$$\rho_j = \mathbb{P}\big[\mathbb{P}[X_i \in \cdot\,|\mathcal{H}_i^j] \in \cdot\,\big] \qquad (i = 1, \ldots, k,\ j = 1, 2), \tag{6.13}$$

and $\mathcal{H}_i^1 \subset \mathcal{H}_i^2$ for all $i = 1, \ldots, k$. Let ω be an independent Ω-valued random variable with law r and let $\mathcal{F}$ be the σ-field generated by ω. Then (6.6) says that

$$\check{T}(\rho_j) = \mathbb{P}\big[\mathbb{P}[\gamma[\omega](X_1, \ldots, X_k) \in \cdot\,|\mathcal{H}_1^j \vee \cdots \vee \mathcal{H}_k^j \vee \mathcal{F}] \in \cdot\,\big] \qquad (j = 1, 2). \tag{6.14}$$

Since $\mathcal{H}_1^1 \vee \cdots \vee \mathcal{H}_k^1 \vee \mathcal{F} \subset \mathcal{H}_1^2 \vee \cdots \vee \mathcal{H}_k^2 \vee \mathcal{F}$, this proves that $\check{T}(\rho_1) \leq_{\rm cv} \check{T}(\rho_2)$.

Let μ be a solution to the RDE (1.4). Then by Lemma 2 for $\rho \in \mathcal{P}(\mathcal{P}(S))_\mu$ we have $\check{T}(\rho)^{(1)} = T(\rho^{(1)}) = T(\mu) = \mu$, proving that $\check{T}$ maps $\mathcal{P}(\mathcal{P}(S))_\mu$ into itself. It will be convenient to combine the proof of the remaining statements of Proposition 3 with the proof of Proposition 4. To check (as claimed in Proposition 4) that $(\omega_{\mathbf{i}}, \xi_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ is an RTP corresponding to the map $\check{\gamma}$ and to $\underline{\mu}$, we need to check that:

- (i) The $(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ are i.i.d.
- (ii) For each $t \geq 1$, the $(\xi_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(t)}}$ are i.i.d. with common law $\underline{\mu}$ and independent of $(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_{(t)}}$.
- (iii) $\xi_{\mathbf{i}} = \check{\gamma}[\omega_{\mathbf{i}}](\xi_{\mathbf{i}1}, \ldots, \xi_{\mathbf{i}\kappa(\omega_{\mathbf{i}})})$ $(\mathbf{i} \in \mathbb{T})$.

Here (i) is immediate. Since $\xi_{\mathbf{i}}$ depends only on $(\omega_{\mathbf{ij}})_{\mathbf{j}\in\mathbb{T}}$, it is also clear that the $(\xi_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(t)}}$ are i.i.d. and independent of $(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_{(t)}}$. To see that their common law is $\underline{\mu}$, we may equivalently show that $\xi_\varnothing$ has law $\underline{\mu}$. Thus, we are left with the task to prove (iii) and

- (iv) $\mathbb{P}[\xi_\varnothing \in \cdot\,] = \underline{\mu}$.

Let $\mathcal{F}^{\mathbf{i}}$ denote the σ-field generated by $(\omega_{\mathbf{ij}})_{\mathbf{j}\in\mathbb{T}}$. Then, for any $\mathbf{i} \in \mathbb{T}$,

$$\xi_{\mathbf{i}} = \mathbb{P}[X_{\mathbf{i}} \in \cdot\,|\mathcal{F}^{\mathbf{i}}] = \mathbb{P}\big[\gamma[\omega_{\mathbf{i}}](X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}\kappa(\omega_{\mathbf{i}})}) \in \cdot\,\big|\mathcal{F}^{\mathbf{i}}\big]. \tag{6.15}$$

Conditional on $\mathcal{F}^{\mathbf{i}}$, the random variables $X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}k_{\mathbf{i}}}$ are independent with respective laws $\xi_{\mathbf{i}1}, \ldots, \xi_{\mathbf{i}k_{\mathbf{i}}}$, and hence $\gamma[\omega_{\mathbf{i}}](X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}\kappa(\omega_{\mathbf{i}})})$ has law $\check{\gamma}[\omega_{\mathbf{i}}](\xi_{\mathbf{i}1}, \ldots, \xi_{\mathbf{i}\kappa(\omega_{\mathbf{i}})})$, proving (iii).

To prove also (iv), we first need to prove (4.9) from Proposition 3. Fix $t \geq 1$ and for $\mathbf{i} \in \mathbb{T}_{(t)} \cup \partial\mathbb{T}_{(t)}$, let $\mathcal{F}_t^{\mathbf{i}}$ denote the σ-field generated by $\{\omega_{\mathbf{ij}} : \mathbf{j} \in \mathbb{T},\ |\mathbf{ij}| < t\}$. In particular, if $\mathbf{i} \in \partial\mathbb{T}_{(t)}$, then $\mathcal{F}_t^{\mathbf{i}}$ is the trivial σ-field. Set

$$\xi_{\mathbf{i}}^t := \mathbb{P}[X_{\mathbf{i}} \in \cdot\,|\mathcal{F}_t^{\mathbf{i}}] \qquad (\mathbf{i} \in \mathbb{T}_{(t)} \cup \partial\mathbb{T}_{(t)}). \tag{6.16}$$

In particular, $\xi^t_{\mathbf{i}} = \mu$ a.s. for $\mathbf{i} \in \partial\mathbb{T}_{(t)}$. Arguing as before, we see that

$$\xi^t_{\mathbf{i}} = \check{\gamma}[\omega_{\mathbf{i}}]\big(\xi^t_{\mathbf{i}1}, \ldots, \xi^t_{\mathbf{i}\kappa(\omega_{\mathbf{i}})}\big) \qquad (\mathbf{i} \in \mathbb{T}_{(t)}), \tag{6.17}$$

and hence

$$\check{T}^t(\delta_\mu) = \mathbb{P}[\xi^t_\varnothing \in \cdot\,]. \tag{6.18}$$

By martingale convergence,

$$\xi^t_\varnothing = \mathbb{P}[X_\varnothing \in \cdot\,|\mathcal{F}^\varnothing_t] \underset{t\to\infty}{\longrightarrow} \mathbb{P}[X_\varnothing \in \cdot\,|(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}] = \xi_\varnothing \quad \text{a.s.} \tag{6.19}$$

Combining this with (6.18), we obtain (4.9) where $\underline{\mu}$ is in fact the law of $\xi_\varnothing$, proving (iv) as well. This completes the proof that $(\omega_{\mathbf{i}}, \xi_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ is an RTP corresponding to the map $\check{\gamma}$ and $\underline{\mu}$.

The proof that $(\omega_{\mathbf{i}}, \delta_{X_{\mathbf{i}}})_{\mathbf{i}\in\mathbb{T}}$ is an RTP corresponding to the map $\check{\gamma}$ and $\overline{\mu}$ is simpler. It is clear that (i) the $(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ are i.i.d., and (ii) for each $t \geq 1$, the $(\delta_{X_{\mathbf{i}}})_{\mathbf{i}\in\partial\mathbb{T}_{(t)}}$ are i.i.d. with common law $\overline{\mu}$ and independent of $(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_{(t)}}$. To prove that also (iii) $\delta_{X_{\mathbf{i}}} = \check{\gamma}[\omega_{\mathbf{i}}]\big(\delta_{X_{\mathbf{i}1}}, \ldots, \delta_{X_{\mathbf{i}\kappa(\omega_{\mathbf{i}})}}\big)$ ($\mathbf{i} \in \mathbb{T}$), it suffices to show that for any measurable $g : S^k \to S$,

$$\check{g}(\delta_{x_1}, \ldots, \delta_{x_k}) = \delta_{g(x_1,\ldots,x_k)}. \tag{6.20}$$

By definition, the left-hand side of this equation is the law of $g(X_1, \ldots, X_k)$, where $X_1, \ldots, X_k$ are independent with laws $\delta_{x_1}, \ldots, \delta_{x_k}$, so the statement is obvious.

This completes the proof of Proposition 4. Moreover, since the marginal law of an RTP solves the corresponding RDE, our proof also shows that the measures $\underline{\mu}$ and $\overline{\mu}$ solve the higher-level RDE (4.3).

In view of this, to complete the proof of Proposition 3, it suffices to prove (4.10). If ρ solves the higher-level RDE (4.3), then applying $\check{T}^t$ to (4.7), using the monotonicity of $\check{T}$ with respect to the convex order, we see that $\check{T}^t(\delta_\mu) \leq_{\rm cv} \rho \leq_{\rm cv} \overline{\mu}$ for all t. Letting $t \to \infty$, (4.10) follows. ∎ ∎

A The convex order

By definition, a *G_δ-set* is a set that is a countable intersection of open sets. By [Bou58, §6 No. 1, Theorem. 1], for a metrizable space S, the following statements are equivalent.

(i) S is Polish.

(ii) There exists a metrizable compactification $\overline{S}$ of S such that S is a G_δ-subset of $\overline{S}$.

(iii) For each metrizable compactification $\overline{S}$ of S, S is a G_δ-subset of $\overline{S}$.

Moreover, a subset $S' \subset S$ of a Polish space S is Polish in the induced topology if and only if S' is a G_δ-subset of S.

Let S be a Polish space. Recall that $\mathcal{P}(S)$ denotes the space of probability measures on S, equipped with the topology of weak convergence. In what follows, we fix a metrizable compactification $\overline{S}$ of S. Then we can identify the space $\mathcal{P}(S)$ (including its topology) with the space of probability measures μ on $\overline{S}$ such that $\mu(S) = 1$. By Prohorov's theorem, $\mathcal{P}(\overline{S})$ is compact, so $\mathcal{P}(\overline{S})$ is a metrizable compactification of $\mathcal{P}(S)$. Recall the definition of $\mathcal{P}(\mathcal{P}(S))_\mu$ from (4.5).

Lemma 9 (Measures with given mean) *For any $\mu \in \mathcal{P}(S)$, the space $\mathcal{P}(\mathcal{P}(S))_\mu$ is compact.*

Proof Since any $\rho \in \mathcal{P}(\mathcal{P}(\overline{S}))$ whose first moment measure is μ must be concentrated on $\mathcal{P}(S)$, we can identify $\mathcal{P}(\mathcal{P}(S))_\mu$ with the space of probability measures on $\mathcal{P}(\overline{S})$ whose first moment measure is μ. From this we see that $\mathcal{P}(\mathcal{P}(S))_\mu$ is a closed subset of $\mathcal{P}(\mathcal{P}(\overline{S}))$ and hence compact. ∎ ∎

We let $\mathcal{C}(\overline{S})$ denote the space of all continuous real functions on $\overline{S}$, equipped with the supremumnorm, and we let $B(\overline{S})$ denote the space of bounded measurable real functions on $\overline{S}$. The following fact is well-known (see, e.g., [Car00, Cor 12.11]).

Lemma 10 (Space of continuous functions) *$\mathcal{C}(\overline{S})$ is a separable Banach space.*

For each $f \in \mathcal{C}(\overline{S})$, we define an affine function $l_f \in \mathcal{C}(\mathcal{P}(\overline{S}))$ by $l_f(\mu) := \int f \,\mathrm{d}\mu$. The following lemma says that all continuous affine functions on $\mathcal{P}(\overline{S})$ are of this form.

Lemma 11 (Continuous affine functions) *A function $\phi \in \mathcal{C}(\mathcal{P}(\overline{S}))$ is affine if and only if $\phi = l_f$ for some $f \in \mathcal{C}(\overline{S})$.*

Proof Let $\phi : \mathcal{P}(\overline{S}) \to \mathbb{R}$ be affine and continuous. Since ϕ is continuous, setting $f(x) := \phi(\delta_x)$ $(x \in \overline{S})$ defines a continuous function $f : \overline{S} \to \mathbb{R}$. Since ϕ is affine, $\phi(\mu) = l_f(\mu)$ whenever μ is a finite convex combination of delta measures. Since such measures are dense in $\mathcal{P}(\overline{S})$ and ϕ is continuous, we conclude that $\phi = l_f$. ∎ ∎

Lemma 12 (Lower semi-continuous convex functions) *Let $C \subset \mathcal{C}(\overline{S})$ be convex, closed, and nonempty. Then*

$$\phi := \sup_{f \in C} l_f \tag{A.1}$$

defines a lower semi-continuous convex function $\phi : \mathcal{P}(\overline{S}) \to (-\infty, \infty]$. Conversely, each such ϕ is of the form (A.1).

Proof It is straightforward to check that (A.1) defines a lower semi-continuous convex function $\phi : \mathcal{P}(\overline{S}) \to (-\infty, \infty]$. To prove that every such function is of the form (A.1), let $\mathcal{C}(\overline{S})'$ denote the dual of the Banach space $\mathcal{C}(\overline{S})$, i.e., $\mathcal{C}(\overline{S})'$ is the space of all continuous linear forms $l : \mathcal{C}(\overline{S}) \to \mathbb{R}$. We equip $\mathcal{C}(\overline{S})'$ with the weak-$*$ topology, i.e., the weakest topology that makes the maps $l \mapsto l(f)$ continuous for all $f \in \mathcal{C}(\overline{S})$. Then $\mathcal{C}(\overline{S})'$ is a locally convex topological vector space and by the Riesz-Markov-Kakutani representation theorem, we can view $\mathcal{P}(\overline{S})$ as a convex compact metrizable subset of $\mathcal{C}(\overline{S})'$. Now any lower semi-continuous convex function $\phi : \mathcal{P}(\overline{S}) \to (-\infty, \infty]$ can be extended to $\mathcal{C}(\overline{S})'$ by putting $\phi := \infty$ on the complement of $\mathcal{P}(\overline{S})$. Applying [CV77, Thm I.3] we obtain that ϕ is the supremum of all continuous affine functions that lie below it. By Lemma 11, we can restrict ourselves to continuous affine functions of the form l_f with $f \in \mathcal{C}(\overline{S})$. It is easy to see that $\{f \in \mathcal{C}(\overline{S}) : l_f \leq \phi\}$ is closed and convex, proving that every lower semi-continuous convex function $\phi : \mathcal{P}(\overline{S}) \to (-\infty, \infty]$ is of the form (A.1). ∎ ∎

We define

$$\mathcal{C}_{\mathrm{cv}}\big(\mathcal{P}(\overline{S})\big) := \big\{\phi \in \mathcal{C}(\mathcal{P}(\overline{S})) : \phi \text{ is convex}\big\} \tag{A.2}$$

If two probability measures $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(S))$ satisfy the equivalent conditions of the following theorem, then we say that they are ordered in the *convex order*, and we denote this as $\rho_1 \leq_{\rm cv} \rho_2$. The fact that $\leq_{\rm cv}$ defines a partial order will be proved in Lemma 15 below. The convex order can be defined more generally for $\rho_1, \rho_2 \in \mathcal{P}(C)$ where C is a convex space, but in the present paper we will only need the case $C = \mathcal{P}(\overline{S})$.

Theorem 13 (The convex order for laws of random probability measures) *Let S be a Polish space and let $\overline{S}$ be a metrizable compactification of S. Then, for $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(S))$, the following statements are equivalent.*

(i) $\displaystyle\int \phi \,\mathrm{d}\rho_1 \leq \int \phi \,\mathrm{d}\rho_2$ *for all* $\phi \in \mathcal{C}_{\rm cv}\big(\mathcal{P}(\overline{S})\big)$.

(ii) *There exists an S-valued random variable X defined on some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and sub-σ-fields $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}$ such that* $\rho_i = \mathbb{P}\big[\mathbb{P}[X \in \cdot\,|\mathcal{F}_i] \in \cdot\,\big]$ $(i = 1, 2)$.

Proof For any probability kernel P on $\mathcal{P}(\overline{S})$, measure $\rho \in \mathcal{P}(\overline{S})$, and function $\phi \in \mathcal{C}(\mathcal{P}(\overline{S}))$, we define $\rho P \in \mathcal{P}(\mathcal{P}(\overline{S}))$ and $P\phi \in B(\mathcal{P}(\overline{S}))$ by

$$\rho P := \int \rho(\mathrm{d}\mu) P(\mu, \,\cdot\,) \quad \text{and} \quad P\phi := \int P(\,\cdot\,, \mathrm{d}\mu)\phi(\mu). \tag{A.3}$$

By definition, a *dilation* is a probability kernel P such that $Pl_f = l_f$ for all $f \in \mathcal{C}(\overline{S})$.

As in the proof of Lemma 12, we can view $\mathcal{P}(\overline{S})$ as a convex compact metrizable subset of the locally convex topological vector space $\mathcal{C}(\overline{S})'$. Then [Str65, Thm 2] tells us that (i) is equivalent to:

(iii) There exists a dilation P on $\mathcal{P}(\overline{S})$ such that $\rho_2 = \rho_1 P$.

To see that this implies (ii), let ξ_1, ξ_2 be $\mathcal{P}(\overline{S})$-valued random variables such that ξ_1 has law ρ_1 and the conditional law of ξ_2 given ξ_1 is given by P. Let $\mathcal{F}_1$ be the σ-field generated by ξ_1, let $\mathcal{F}_2$ be the σ-field generated by (ξ_1, ξ_2), and let X be an $\overline{S}$-valued random variable whose conditional law given $\mathcal{F}_2$ is given by ξ_2. Then

$$\mathbb{P}\big[\mathbb{P}[X \in \cdot\,|\mathcal{F}_2] \in \cdot\,\big] = \mathbb{P}[\xi_2 \in \cdot\,] = \rho_1 P = \rho_2. \tag{A.4}$$

Since P is a dilation

$$\mathbb{E}[f(X)\,|\,\mathcal{F}_1] = \mathbb{E}\big[\mathbb{E}[f(X)\,|\,\mathcal{F}_2]\,\big|\,\mathcal{F}_1\big] = \mathbb{E}\big[l_f(\xi_2)\,\big|\,\mathcal{F}_1\big] = \int P(\xi_1, \mathrm{d}\mu) l_f(\mu) = l_f(\xi_1) \tag{A.5}$$

for all $f \in \mathcal{C}(\overline{S})$, and hence

$$\mathbb{P}\big[\mathbb{P}[X \in \cdot\,|\mathcal{F}_1] \in \cdot\,\big] = \mathbb{P}\big[\xi_1 \in \cdot\,\big] = \rho_1. \tag{A.6}$$

We note that since $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(S))$, we have $\xi_1, \xi_2 \in \mathcal{P}(S)$ a.s. and hence $X \in S$ a.s. This proves the implication (iii)$\Rightarrow$(ii).

To complete the proof, it suffices to show that (ii)$\Rightarrow$(i). By Lemma 12, each $\phi \in \mathcal{C}_{\rm cv}(\mathcal{P}(\overline{S}))$ is of the form $\phi = \sup_{f \in C} l_f$ for some $C \subset \mathcal{C}(\overline{S})$. Then (ii) implies

$$\begin{aligned}\int \phi \,\mathrm{d}\rho_1 &= \mathbb{E}\big[\sup_{f \in C} \mathbb{E}[f(X)\,|\,\mathcal{F}_1]\big] = \mathbb{E}\big[\sup_{f \in C} \mathbb{E}\big[\mathbb{E}[f(X)\,|\,\mathcal{F}_2]\,\big|\,\mathcal{F}_1\big]\big] \\ &\leq \mathbb{E}\big[\mathbb{E}\big[\sup_{f \in C} \mathbb{E}[f(X)\,|\,\mathcal{F}_2]\,\big|\,\mathcal{F}_1\big]\big] = \mathbb{E}\big[\sup_{f \in C} \mathbb{E}[f(X)\,|\,\mathcal{F}_2]\big] = \int \phi \,\mathrm{d}\rho_2.\end{aligned} \tag{A.7}$$

∎ ∎

The n-th moment measure $\rho^{(n)}$ associated with a probability law $\rho \in \mathcal{P}(\mathcal{P}(\overline{S}))$ has been defined in (4.1). The following lemma links the first and second moment measures to the convex order.

Lemma 14 (First and second moment measures) *Let S be a Polish space. Assume that $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(S))$ satisfy $\rho_1 \leq_{\rm cv} \rho_2$. Then $\rho_1^{(1)} = \rho_2^{(1)}$ and*

$$\int \rho_1^{(2)}(\mathrm{d}x, \mathrm{d}y) f(x) f(y) \leq \int \rho_2^{(2)}(\mathrm{d}x, \mathrm{d}y) f(x) f(y) \qquad \big(f \in B(S)\big). \tag{A.8}$$

If $\rho_1 \leq_{\rm cv} \rho_2$ and (A.8) holds with equality for all bounded continuous $f : S \to \mathbb{R}$, then $\rho_1 = \rho_2$.

Proof By Theorem 13, there exists an $\overline{S}$-valued random variable X defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and sub-σ-fields $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}$ such that $\rho_i = \mathbb{P}\big[\mathbb{P}[X \in \cdot\,|\mathcal{F}_i] \in \cdot\,\big]$ $(i = 1, 2)$. Since for each $f \in B(S)$

$$\int \rho_1^{(1)}(\mathrm{d}x) f(x) = \mathbb{E}\big[\mathbb{E}[f(X)\,|\mathcal{F}_1]\big] = \mathbb{E}[f(X)] = \mathbb{E}\big[\mathbb{E}[f(X)\,|\mathcal{F}_2]\big] = \int \rho_2^{(1)}(\mathrm{d}x) f(x), \tag{A.9}$$

we see that $\rho_1^{(1)} = \rho_2^{(1)}$. Fix $f \in B(S)$ and set $M_i := \mathbb{E}[f(X)\,|\mathcal{F}_i]$ $(i = 1, 2)$. Then

$$\begin{aligned} &\int \rho_2^{(2)}(\mathrm{d}x, \mathrm{d}y) f(x) f(y) = \mathbb{E}\big[\mathbb{E}[f(X)\,|\mathcal{F}_2]^2\big] = \mathbb{E}[M_2^2] \\ &\quad = \mathbb{E}[M_1^2] + \mathbb{E}\big[(M_2 - M_1)^2\big] \geq \mathbb{E}[M_1^2] = \int \rho_1^{(2)}(\mathrm{d}x, \mathrm{d}y) f(x) f(y), \end{aligned} \tag{A.10}$$

proving (A.8). Let $\overline{S}$ be a metrizable compactification of S. If $\rho_1 \leq_{\rm cv} \rho_2$ and (A.8) holds with equality for all bounded continuous $f : S \to \mathbb{R}$, then (A.10) tells us that $M_1 = M_2$ for each $f \in \mathcal{C}(\overline{S})$, i.e.,

$$\mathbb{E}[f(X)\,|\mathcal{F}_1] = \mathbb{E}[f(X)\,|\mathcal{F}_2] \text{ a.s. for each } f \in \mathcal{C}(\overline{S}). \tag{A.11}$$

By Lemma 10, we can choose a countable dense set $\mathcal{D} \subset \mathcal{C}(\overline{S})$. Then $\mathbb{E}[f(X)\,|\mathcal{F}_1] = \mathbb{E}[f(X)\,|\mathcal{F}_2]$ for all $f \in \mathcal{D}$ a.s. and hence $\mathbb{P}[X \in \cdot\,|\mathcal{F}_1] = \mathbb{P}[X \in \cdot\,|\mathcal{F}_2]$ a.s., proving that $\rho_1 = \rho_2$. ∎ ∎

The following lemma shows that the convex order is a partial order,

Lemma 15 (Convex functions are distribution determining) *If $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(\overline{S}))$ satisfy $\int \phi \,\mathrm{d}\rho_1 = \int \phi \,\mathrm{d}\rho_2$ for all $\phi \in \mathcal{C}_{\rm cv}(\mathcal{P}(\overline{S}))$, then $\rho_1 = \rho_2$.*

Proof For any $f \in \mathcal{C}(\overline{S})$ and $\rho \in \mathcal{P}(\mathcal{P}(\overline{S}))$,

$$\begin{aligned} &\int_{\overline{S}^2} \rho^{(2)}(\mathrm{d}x, \mathrm{d}y) f(x) f(y) \\ &\quad = \int_{\mathcal{P}(\overline{S})} \rho(\mathrm{d}\mu) \int_{\overline{S}^2} \mu(\mathrm{d}x)\mu(\mathrm{d}y) f(x) f(y) = \int_{\mathcal{P}(\overline{S})} \rho(\mathrm{d}\mu) l_f(\mu)^2. \end{aligned} \tag{A.12}$$

Therefore, since l_f^2 is a convex function, $\int \phi \,\mathrm{d}\rho_1 = \int \phi \,\mathrm{d}\rho_2$ for all $\phi \in \mathcal{C}_{\rm cv}(\mathcal{P}(\overline{S}))$ implies equality in (A.8) and hence, by Lemma 14, $\rho_1 = \rho_2$. ∎ ∎

B Open Problem 12 of Aldous and Bandyopadhyay

We have seen that the use of the higer-level map from Section 4 and properties of the convex order lead to an elegant and short proof of Theorem 1, which is similar to [AB05, Thm 11]. The most significant improvement over [AB05, Thm 11] is that the implication (ii)⇒(i) is shown without a continuity assumption on the map T, solving Open Problem 12 of [AB05]. If one is only interested in solving this open problem, taking the proof of [AB05, Thm 11] for granted, then it is possible to give a shorter argument that does not involve the higer-level map and the convex order.

One way to prove the implication (ii)⇒(i) in Theorem 1 is to show that nonendogeny implies the existence of a measure $\nu \in \mathcal{P}(S^2)_\mu$ such that $T^{(2)}(\nu) = \nu$ and $\nu \neq \overline{\mu}^{(2)}$. In [AB05], such a ν was constructed as the weak limit of measures ν_n which satisfied $T^{(2)}(\nu_n) = \nu_{n+1}$; however, to conclude that $T^{(2)}(\nu) = \nu$ they then needed to assume the continuity of $T^{(2)}$. Their Open Problem 12 asks if this continuity assumption can be removed.

In our proof of Theorem 1, we take $\nu = \underline{\mu}^{(2)}$, which by Theorem 5 and Lemma 14 from Appendix A satisfies $\nu \neq \overline{\mu}^{(2)}$ if and only if the RTP corresponding to μ is not endogenous, and by Lemma 4.4 satisfies $T^{(2)}(\nu) = \nu$.

Antar Bandyopadhyay told us that shortly after the publication of [AB05], he learned that their Open Problem 12 could be solved by adapting the proof of the implication (3)⇒(2) of [BL07, Théorème 9] to the setting of RTPs. To the best of our knowledge, this observation has not been published. The setting of [BL07, Théorème 9] are positive recurrent Markov chains with countable state space, which are a very special case of the RTPs we consider. In view of this, we sketch their argument here in our general setting and show how it relates to our argument.

Let $(\omega_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be an RTP corresponding to the map γ and a solution μ of a RDE. Construct $(Y_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ such that $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ and $(Y_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ are conditionally independent and identically distributed given $(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$. Then $X_\varnothing = Y_\varnothing$ a.s. if and only if the RTP corresponding to μ is endogenous. Let ν denote the law of $(X_\varnothing, Y_\varnothing)$. Then $\nu = \overline{\mu}^{(2)}$ if and only if endogeny holds. In view of this, to prove the implication (ii)⇒(i) in Theorem 1, it suffices to show that ν solves the bivariate RDE $T^{(2)}(\nu) = \nu$. This will follow provided we show that

$$\big(\omega_{\mathbf{i}}, (X_{\mathbf{i}}, Y_{\mathbf{i}})\big)_{\mathbf{i}\in\mathbb{T}} \tag{B.1}$$

is an RTP corresponding to the map $\gamma^{(2)}$ and ν, i.e.,

$$\begin{array}{rl}
\text{(i)} & \text{the } (\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}} \text{ are i.i.d.,} \\
\text{(ii)} & \text{for each } t \geq 1, \text{ the } (X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(t)}} \text{ are i.i.d. with common law } \nu \\
& \text{and independent of } (\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_{(t)}}, \\
\text{(iii)} & (X_{\mathbf{i}}, Y_{\mathbf{i}}) = \gamma^{(2)}[\omega_{\mathbf{i}}]\big((X_{\mathbf{i}1}, Y_{\mathbf{i}1}), \ldots, (X_{\mathbf{i}k_{\mathbf{i}}}, Y_{\mathbf{i}\kappa(\omega_{\mathbf{i}})})\big) \qquad (\mathbf{i} \in \mathbb{T}).
\end{array} \tag{B.2}$$

Here (i) and (iii) are trivial. To prove property (ii), set

$$\Lambda_{\mathbf{i}} := \big(X_{\mathbf{i}}, (\omega_{\mathbf{ij}})_{\mathbf{j}\in\mathbb{T}}\big) \quad \text{and} \quad \Lambda^{(2)}_{\mathbf{i}} := \big(X_{\mathbf{i}}, Y_{\mathbf{i}}, (\omega_{\mathbf{ij}})_{\mathbf{j}\in\mathbb{T}}\big) \qquad (\mathbf{i} \in \mathbb{T}). \tag{B.3}$$

Then the $(\Lambda_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ are identically distributed. Moreover, for each $t \geq 1$, the $(\Lambda_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(t)}}$ are independent of each other and of $(\omega_{\mathbf{k}})_{\mathbf{k}\in\mathbb{T}_{(t)}}$. Recall that $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ and $(Y_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ are conditionally independent and identically distributed given $(\omega_{\mathbf{k}})_{\mathbf{k}\in\mathbb{T}}$. Since the conditional law of $X_{\mathbf{i}}$ given $(\omega_{\mathbf{k}})_{\mathbf{k}\in\mathbb{T}}$ only depends on $(\omega_{\mathbf{ij}})_{\mathbf{j}\in\mathbb{T}}$, the same is true for $Y_{\mathbf{i}}$. Using this, it is not hard to see that the $(\Lambda^{(2)}_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ are identically distributed and for each $t \geq 1$, the $(\Lambda^{(2)}_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(t)}}$ are independent of each other and of $(\omega_{\mathbf{k}})_{\mathbf{k}\in\mathbb{T}_{(t)}}$, and this in turn implies (ii).

In fact, since the law of $(X_\varnothing, Y_\varnothing)$ is the second moment measure of the random measure $\xi_\varnothing$ from Proposition 4, the measure ν constructed here is the same as our measure $\underline{\mu}^{(2)}$. Thus, our argument and the one from [BL07] are both based on the same solution of the bivariate RDE.

Acknowledgements

We thank Wolfgang Löhr and Jan Seidler for their help with Lemmas 11 and 12, respectively. We thank David Aldous, Antar Bandyopadhyay, and Christophe Leuridan for answering our questions about their work.

References

[AB05] D.J. Aldous and A. Bandyopadhyay. A survey of max-type recursive distributional equations. *Ann. Appl. Probab.* 15(2) (2005), 1047–1110.

[Ant06] A. Bandyopadhyay. A necessary and sufficient condition for the tail-triviality of a recursive tree process. *Sankhya* 68(1) (2006), 1–23.

[BL07] J. Brossard and C. Leuridan. Chaînes de Markov consructives indexées par $\mathbb{Z}$. *Ann. Probab.* 35(2) (2007), 715–731.

[Bou58] N. Bourbaki. *Éléments de Mathématique. VIII. Part. 1: Les Structures Fondamentales de l'Analyse. Livre III: Topologie Générale. Chap. 9: Utilisation des Nombres Réels en Topologie Générale. 2iéme éd.* Actualités Scientifiques et Industrielles 1045. Hermann & Cie, Paris, 1958.

[Car00] N.L. Carothers. *Real Analysis.* Cambridge University Press, 2000.

[CV77] C. Castaing and M. Valadier. *Convex Analysis and Measurable Multifunctions.* Springer, Berlin, 1977.

[Geo11] H.-O. Georgii. *Gibbs measures and phase transitions. 2nd ed.* De Gruyter, Berlin, 2011.

[Mos01] E. Mossel. Reconstruction on trees: beating the second eigenvalue. *Ann. Appl. Probab.* 11(1) (2001), 285–300.

[PW96] J.G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* 9 (1996), 223–252.

[Ros59] M. Rosenblatt. Stationary processes as shifts of functions of independent random variables. *J. Math. Mech.* 8 (1959), 665–681.

[Str65] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Stat.* 36 (1965), 423–439.