A CHARACTERISATION OF COMPATIBLE STATE SPACE AGGREGATIONS FOR DISCRETE DYNAMICAL SYSTEMS

CHRIS GOOD, DAVID PARKER, MATE PULJIZ, AND JONATHAN E. ROWE

Abstract. We study coarse grainings of discrete dynamical systems on $\mathbb{R}^n$. In particular we are interested in state space aggregations. We give a complete characterisation of compatible aggregations for analytic maps on $\mathbb{R}^n$. We show how our results apply to artificial chemistries, Random Heuristic Search models, and related finite population models.

1. Introduction

Modelling real life dynamical systems is a challenging problem. Whether we are exploring social dynamics on a large scale, or trying to understand biological interactions at the microscopic level of a cell, the number of variables involved to describe these processes is often too large for any meaningful computer simulation or analysis to take place in a reasonable amount of time. Researchers then usually resort to various techniques of simplifying the underlying model, thus reducing its dimensionality and enabling the computation of model predictions.

In this paper we look into a systematic approach for finding these simplifications for a class of models stemming from the Random Heuristic Search framework developed by Vose in [20]. Let us explain on a toy model the kind of systems we have in mind (see also Example 4 below). Imagine a country in which each person favours one out of 3 political options, A, B, or C. In interactions with other people, each pair of individuals will reach a consensus on which party they would like to vote for. Furthermore, every person preferring A will convert any supporter of B or C, and likewise a supporter of B will convert supporters of C. Let $T(p_{\mathrm{A}}, p_{\mathrm{B}}, p_{\mathrm{C}}) = (r_{\mathrm{A}}, r_{\mathrm{B}}, r_{\mathrm{C}})$ be a map (later called a heuristic) that encodes this dynamics in the following way. Given a population with $100 \cdot p_{\mathrm{X}}\%$ supporters of the option X, r_{Y} is the probability that two individuals chosen at random from that population adopt the common view Y. In fact, in order to keep track of the number of supporters of A it is not necessary to distinguish supporters of B and C. They can all be lumped together, and we will have simplified our model and reduced its dimensionality by 1.

Intuitively, *coarse graining* (also *(ordinary) lumping* or *aggregation*) refers to clustering portions of the state space in such a way that the dynamics on these 'higher level' states is still well-defined. In particular, we are interested in state space aggregations that correspond to the changes of variables of the form $X_i = x_{i_1} + \cdots + x_{i_k}$ where each of the variables x_j occurs exactly once in one X_i, thus effectively partitioning the set of variables. Coarse graining has been previously studied by Vose and his collaborators and some useful criteria have been devised that guarantee existence of coarse grainings in certain cases ([2, 15–17]).

Other authors, most prominently Rabitz and collaborators (see [10–12]), and more recently Jacobi ([7]), and Tomlin et al. ([19]) explored the coarse grainings of continuous systems given by differential equations. It turns out that the theory of linear coarse grainings coincides in both the discrete and continuous systems.

We would like to stress that there are other approaches to simplifying the dynamics of iterated dynamical systems. These model reduction techniques include, most notably, the singular value decomposition and variants thereof. They, however, will not be a subject of our interest in this study, since they usually do not preserve the inner dynamics of the system which is our primary concern. For some papers that do follow this line of research see [6, 8].

This paper is organised as follows. In Section 2 we set up the notation, give basic definitions and survey known results. An emphasis is put on continuous-time systems and how they come together with our discrete-time theory. We also exemplify some of their differences. In Section 3 a few motivating examples are given and special instances of the main theorem are derived. These already generalise some of the results from [17] and hint at the general framework that is developed throughout Sections 4 and 5, and which incorporates, amongst others, the model given at the beginning.

We start with an idealised, infinite population dynamics given by a heuristic map. Then for any $r \in \mathbb{N}$, which represents the real size of the population, Vose showed how to derive various associated Markov chains modelling the changes this finite population undergoes. In [20] he proved that the aggregations of the heuristic map naturally extend to aggregations of these derived models. The workings of this machinery are, for completeness, explained in Section 4. Additionally, in Appendix A we prove that this applies to some other classes of derived models reinforcing the view that it is only aggregations of the heuristic that matter. In Section 5 we prove our main theorem, a characterisation of compatible lumpings in terms of the coefficients of the Taylor series of the heuristic map (Theorem 16). This is then applied to systems modelling artificial chemistries.

Concluding remarks and a discussion outlining directions for further research are left for Section 6 and the appendix.

2. General theory

We start by defining a dynamical system, borrowing notation from [18].

Definition 1. Let J be $\mathbb{R} \cap [0, \infty)$ or $\mathbb{Z} \cap [0, \infty)$ and X a non-empty set (state space). A map $\Phi\colon J \times X \to X$ is called a (global autonomous) semi-flow if

(1) $\Phi(0, x) = x$ for all $x \in X$,
(2) $\Phi(t + s, x) = \Phi(t, \Phi(s, x))$ for all $t, s \in J$ and $x \in X$.

Often the space X comes with a topology and then we will require that Φ is a continuous map. To simplify the notation we write $\Phi_t(x)$ instead of $\Phi(t, x)$.

We are interested in conglomerating points of the state space to get, possibly, a simpler system. The following definition captures the idea.

Definition 2. Let (X, Φ) and (Y, Ψ) be two dynamical systems with the same underlying time set J. A (continuous) map $\Xi\colon X \to Y$ is a *coarse graining* if the following diagram commutes for all $t \in J$,

$$\begin{array}{ccc} X & \xrightarrow{\Phi_t} & X \\ \Xi \downarrow & & \downarrow \Xi \\ Y & \xrightarrow{\Psi_t} & Y \end{array}$$

or symbolically $\Xi(\Phi_t(x)) = \Psi_t(\Xi(x))$ is satisfied for all $t \in J$ and $x \in X$.

In the discrete case, a semi-flow is completely determined by the map $\Phi_1 \colon X \to X$, which we denoted by T from now on, and for all $n \in \mathbb{N}$ we have

$$\Phi_n(x) = (\underbrace{\Phi_1 \circ \Phi_1 \circ \cdots \circ \Phi_1}_{n \text{ times}})(x) = (\Phi_1)^n(x) = T^n(x),$$

so it is enough to check the commutativity condition for $t = 1$, i.e.

$$\Xi(T(x)) = \Psi_1(\Xi(x)), \text{ for all } x \in X \tag{1}$$

which implies that coarse graining coincides with the notion of semi-conjugacy in the theory of dynamical systems (see e.g. [1]). For this reason the map Ψ_1 will be denoted by $\tilde{T}$ in the discrete case.

In the continuous case, the state space is usually a differential manifold and the semi-flow is implicitly given as a solution to an ordinary differential equation

$$\dot{\boldsymbol{x}} = A(\boldsymbol{x}). \tag{2}$$

In order to make the dynamics well defined we require that the equation has a unique forward solution. Under a mild assumption of Lipschitz continuity of A, this is guaranteed by the Picard–Lindelöf theorem. In that case, the map $\left[\frac{d}{dt}\Phi_t\right]_{t=t_0} : X \to TX$[1] does not depend on t_0 and defines the vector field of velocities of trajectories on X. If one wishes to investigate the stability of the system, it is not uncommon to fix an equilibrium point and require $A(0) = 0$ as in [19]. We shall not make this assumption. In what follows we do however require that A is at least twice continuously differentiable and refer to these as *smooth systems*. The following lemma was proven in [7] (compare also with [19, Theorem 1.2]).

Lemma 3 ([7], Lemmas 1 and 2)**.** *Let (X, Φ) and (Y, Ψ) be smooth systems given by their respective differential equations*

$$\dot{\boldsymbol{x}} = A(\boldsymbol{x}), \quad \dot{\boldsymbol{y}} = B(\boldsymbol{y}). \tag{3}$$

A smooth map $\Xi \colon X \to Y$ is a coarse graining if and only if

$$(D\Xi)_{\boldsymbol{x}} \cdot A(\boldsymbol{x}) = B(\Xi(\boldsymbol{x})), \text{ for } \boldsymbol{x} \in X \tag{4}$$

where $(D\Xi)_{\boldsymbol{x}}$ denotes the differential of Ξ calculated at the point $\boldsymbol{x}$.

The proof essentially consists of noticing that, when differentiated with respect to t and evaluated at, say, $t = 0$, the condition $\Xi(\Phi_t(x)) = \Psi_t(\Xi(x))$ becomes (4). Note that in the case where Ξ is a linear map[2] this condition (4) reads

$$\Xi(A(\boldsymbol{x})) = B(\Xi(\boldsymbol{x})), \quad \boldsymbol{x} \in X.$$

[1]As is common in the literature on differential geometry TX denotes tangent bundle of X.

[2]This requires the underlying space to be linear, but most often we will be interested in maps on (subsets) of $\mathbb{R}^n$.

Comparing the latter with (1) it is apparent that understanding *linear* coarse grainings of smooth and discrete systems is essentially the same problem. This is perhaps best seen in the example of discrete- and continuous-time Markov chains. Recall that a time-homogeneous discrete-time Markov chain (DTMC) on a finite state space Ω, $|\Omega| = n$, is (probabilistically) completely determined with its initial distribution $\boldsymbol{p} \in \Lambda_n$ and a transition matrix M where $\Lambda_n = \{(p_1, \ldots, p_n)^T \in \mathbb{R}^n_{\geq 0} | \sum_{i=1}^n p_i = 1\}$ and M is an $n \times n$ stochastic matrix, i.e. has non-zero entries and the entries in each column add up to one (see e.g. [13]). In our setting, we can regard this as a system (Λ_n, Φ) on the simplex given by $\Phi_k(\boldsymbol{p}) = M^k \cdot \boldsymbol{p}$, for all $k \in \mathbb{N}_0$.

A continuous-time analogue of DTMCs are continuous-time Markov chains (CTMCs) with a finite state space. The most important class of CTMCs can be constructed as follows. Start with an $n \times n$ matrix $Q = [q_{ij}]$ such that

- $q_{ii} \leq 0$, for all i;
- $q_{ij} \geq 0$, for all $i \neq j$;
- $\sum_{i=1}^n q_{ij} = 0$, for all j.

Set $P(t) = e^{tQ}$. It can be shown that $P(t)$ is stochastic for all $t \in \mathbb{R}_{\geq 0}$. Using the matrix Q as a *generator matrix* it is possible to construct a continuous-time random process $(X_t : t \geq 0)$ such that

$$\begin{aligned}\mathbb{P}(X_{t_{n+1}} = i_{n+1} | X_{t_n} = i_n, \ldots, X_{t_0} = i_0) = \\ = \mathbb{P}(X_{t_{n+1}} = i_{n+1} | X_{t_n} = i_n) = p_{i_{n+1} i_n}(t_{n+1} - t_n),\end{aligned}$$

where the matrix $P(t_{n+1} - t_n)$ has entries $p_{ij}(t_{n+1} - t_n)$. This means that the process that has an initial distribution $\boldsymbol{p}$ over states in Ω, at the time t is distributed as $P(t) \cdot \boldsymbol{p} = e^{tQ} \cdot \boldsymbol{p}$. For more details see Chapters 2 and 3 of [13]. Hence, we can regard such a CTMC as a dynamical system (Λ_n, Φ) given by $\Phi_t(\boldsymbol{p}) = e^{tQ} \cdot \boldsymbol{p}$, for $t \in \mathbb{R}_{\geq 0}$. Note that Φ satisfies the differential equation $\frac{d}{dt}\Phi_t(\boldsymbol{p}) = Q \cdot \Phi_t(\boldsymbol{p})$ and according to what was said above, the linear coarse grainings of that map can be identified as the linear coarse grainings of the associated map $\boldsymbol{x} \mapsto Q \cdot \boldsymbol{x}$.

When we move away from linear coarse grainings things change drastically. But before we discuss what happens in that case we recall a result from [17] that gives necessary and sufficient conditions for a map to be a coarse graining in the discrete-time case.

Theorem 4 ([17], Theorem 1)**.** *Let V and W be open sets in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively and let T be a continuously differentiable function on V that generates a discrete system. Let $\Xi \colon V \to W$ be a smooth map such that its level sets are connected by smooth paths. Then Ξ is a coarse graining of the system T if and only if*

$$(DT)_{\boldsymbol{x}} \cdot T_{\boldsymbol{x}} \subseteq \ker (D\Xi)_{T(\boldsymbol{x})}, \text{ for all } \boldsymbol{x} \in V \tag{5}$$

where $T_{\boldsymbol{x}} \subseteq \ker (D\Xi)_{\boldsymbol{x}}$ is a tangent space at $\boldsymbol{x}$ defined as a linear span of a set of all velocities realised by smooth paths passing through $\boldsymbol{x}$ and attaining values within the same level set of Ξ[3]*.*

The intuition underlying condition (5) is as follows. Given a map Ξ, to check that it is a valid coarse graining we would normally check that any two points belonging to the same class map to points from the same class. Instead we check that classes are mapped one to another locally. The fact that any two points

[3]This is essentially the tangent space at $\boldsymbol{x}$ of the manifold $\{\boldsymbol{y} \in V : \Xi(\boldsymbol{y}) = \Xi(\boldsymbol{x})\}$.

in the same class are connected by a smooth path allows us to extend this local compatibility globally. As the following example shows, the assumption that level sets are smooth-path-connected is crucial for the validity of the result.

Example 1. Let $V = W = \mathbb{R}$, $T(x) = e^x$ and $\Xi(y) = y^2$. It is clear that T_x is the null-space for every $x \in \mathbb{R}$ and that condition (5) is trivially satisfied. Still Ξ is not a valid coarse graining of the system as, for example, $\Xi(1) = \Xi(-1)$ but $\Xi(T(1)) = e^2 \neq e^{-2} = \Xi(T(-1))$.

If Ξ in Theorem 4 is assumed to be linear then the result reduces to Theorem 1 in [15] and condition (5) simplifies to

$$(DT)_{\boldsymbol{x}} \cdot \ker \Xi \subseteq \ker \Xi. \tag{6}$$

Combining these remarks with Theorem 4 and Lemma 3 we get

Corollary 5 ([7], Proposition 1)**.** *Let V be a convex open set in $\mathbb{R}^n$ and (V, Φ) a smooth system given by the differential equation $\dot{\boldsymbol{x}} = A(\boldsymbol{x})$. A linear map $\Xi\colon V \to \mathbb{R}^m$ is a coarse graining if and only if $(DA)_{\boldsymbol{x}} \cdot \ker \Xi \subseteq \ker \Xi$, for all $\boldsymbol{x} \in V$. In this case, the coarse grained system satisfies the equation $\dot{\boldsymbol{y}} = B(\boldsymbol{y})$ where $B\colon \Xi(V) \to \mathbb{R}^m$ is the map given by $B(\Xi(x)) = \Xi(A(x))$.*

The advantage of Theorem 4 over Lemma 3 is that the former does not mention the coarse grained map Ψ in the statement, although, once it is established that Ξ is a coarse graining, the simplified, coarse grained map is easily obtained. In reality, this form is more applicable when we are required to produce a coarse graining for a given map. A similar improvement can be achieved in the continuous setting as well and is essentially the content of [19, Theorems 1.5 and 1.6], but as our setting is slightly different we give a proof below.

Theorem 6. *Let V and W be open sets in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively and let (V, Φ) be a smooth system on V satisfying the differential equation $\dot{\boldsymbol{x}} = A(\boldsymbol{x})$. Let $\Xi\colon V \to W$ be a smooth map such that its level sets are connected by smooth paths. Then Ξ is a coarse graining of the system Φ if and only if for all $\boldsymbol{x} \in V$*

$$(D^2\Xi)_{\boldsymbol{x}}[A(\boldsymbol{x}), \boldsymbol{w}] + (D\Xi)_{\boldsymbol{x}} \cdot (DA)_{\boldsymbol{x}} \cdot \boldsymbol{w} = 0, \text{ for all } \boldsymbol{w} \in T_{\boldsymbol{x}}, \tag{7}$$

where $T_{\boldsymbol{x}} \subseteq \ker (D\Xi)_{\boldsymbol{x}}$ is a set of all velocities realised by smooth paths passing through $\boldsymbol{x}$ and attaining values within the same level set of Ξ as before.

Proof. The proof is very similar to the proof of Theorem 4 in [15]. Let $\boldsymbol{x} \in V$ and $\boldsymbol{w} \neq 0$ a vector in $T_{\boldsymbol{x}} \subseteq \ker(D\Xi)_{\boldsymbol{x}}$. There is a smooth path $\boldsymbol{x}(t)$ in V passing through $\boldsymbol{x}$ having velocity $\boldsymbol{w}$ at $t = 0$ and staying in the same level set of Ξ. According to Lemma 3 $(D\Xi)_{\boldsymbol{x}(t)} \cdot A(\boldsymbol{x}(t))$ is constant on some open interval around zero and hence

$$0 = \frac{d}{dt}\left[(D\Xi)_{\boldsymbol{x}(t)} \cdot A(\boldsymbol{x}(t))\right]_{t=0} = (D^2\Xi)_{\boldsymbol{x}}[A(\boldsymbol{x}), \boldsymbol{w}] + (D\Xi)_{\boldsymbol{x}} \cdot (DA)_{\boldsymbol{x}} \cdot \boldsymbol{w}.$$

Conversely, assume (7) holds. We need to show that the higher level dynamics is well defined. Let $\boldsymbol{v}$ and $\boldsymbol{w}$ be two points from the same level set of Ξ and let $\rho\colon [0, 1] \to V$ be a smooth path (we denote the temporal variable by s) within the same level set of Ξ with $\rho(0) = \boldsymbol{v}$ and $\rho(1) = \boldsymbol{w}$. So it suffices to show that $\Phi_t \circ \rho$ is again a smooth path within the same level set of Ξ for all t. Hence we need to show that for all $t \geq 0$ and $s_0 \in (0, 1)$

$$(D\Xi)_{\Phi_t\rho(s_0)} \cdot \frac{d}{ds}\left[\Phi_t(\rho(s))\right]_{s=s_0} = 0. \tag{8}$$

Here we use the fact that a smooth path stays within the same level set of Ξ if and only if its velocity is normal to $D\Xi$ at all times. Note that (8) is satisfied for $t = 0$ as ρ is indeed a path within a level set of Ξ. So it turns out that it is enough to show that the left side in (8) stays constant as t varies or equivalently

$$\frac{d}{dt}\left[(D\Xi)_{\Phi_t\rho(s_0)} \cdot \frac{d}{ds}\left[\Phi_t(\rho(s))\right]_{s=s_0}\right] = 0. \tag{9}$$

We calculate

$$\begin{aligned}
&\frac{d}{dt}\left[(D\Xi)_{\Phi_t\rho(s_0)} \cdot \frac{d}{ds}\left[\Phi_t(\rho(s))\right]_{s=s_0}\right] = \\
&\quad = (D^2\Xi)_{\Phi_t\rho(s_0)}\left[\frac{d}{ds}\left[\Phi_t(\rho(s))\right]_{s=s_0}, \frac{d}{dt}\left[\Phi_t(\rho(s_0))\right]\right] + (D\Xi)_{\Phi_t\rho(s_0)} \cdot \frac{d}{dt}\frac{d}{ds}\left[\Phi_t(\rho(s))\right]_{s=s_0} = \\
&\quad = (D^2\Xi)_{\boldsymbol{x}(t)}\left[\boldsymbol{w}(t), \dot{\boldsymbol{x}}(t)\right] + (D\Xi)_{\boldsymbol{x}(t)} \cdot \frac{d}{ds}\left[\frac{d}{dt}\Phi_t(\rho(s))\right]_{s=s_0} = \\
&\quad = (D^2\Xi)_{\boldsymbol{x}(t)}\left[\boldsymbol{w}(t), A(\boldsymbol{x}(t))\right] + (D\Xi)_{\boldsymbol{x}(t)} \cdot \frac{d}{ds}\left[A(\Phi_t(\rho(s))\right]_{s=s_0} = \\
&\quad = (D^2\Xi)_{\boldsymbol{x}(t)}\left[\boldsymbol{w}(t), A(\boldsymbol{x}(t))\right] + (D\Xi)_{\boldsymbol{x}(t)} \cdot (DA)_{\boldsymbol{x}(t)} \cdot \frac{d}{ds}\left[\Phi_t(\rho(s))\right]_{s=s_0} = \\
&\quad = (D^2\Xi)_{\boldsymbol{x}(t)}\left[\boldsymbol{w}(t), A(\boldsymbol{x}(t))\right] + (D\Xi)_{\boldsymbol{x}(t)} \cdot (DA)_{\boldsymbol{x}(t)} \cdot \boldsymbol{w}(t)
\end{aligned}$$

where we set $\boldsymbol{x}(t) = \Phi_t\rho(s_0)$, $\boldsymbol{w}(t) = \frac{d}{ds}\left[\Phi_t(\rho(s))\right]_{s=s_0}$ and $\frac{d}{dt}\Phi_t(\rho(s)) = A(\Phi_t\rho(s))$ since the solutions to the system Φ satisfy equation $\dot{\boldsymbol{x}} = A(\boldsymbol{x})$. Finally, we note that the expression above, under the assumption (7), equals zero as required. □

The improvement made by this theorem is somewhat artificial as solving the differential equation above is by no means an easy task. However, imposing additional assumptions on coarse graining can possibly simplify it. For example it is obvious that assuming Ξ is linear the statement of the theorem reduces to Corollary 5.

Example 2. Observe the continuous-time system in the plane $\mathbb{R}^2$ given by

$$\dot{x} = x, \qquad \dot{y} = 2y.$$

Using Theorem 6 we can show that $\Xi\colon \mathbb{R}^2 \to \mathbb{R}$, $\Xi(x, y) = xy$ coarse grains the initial system to the system given by $\dot{z} = 3z$. We calculate

$$(D\Xi)_{(x,y)} = \begin{bmatrix} y & x \end{bmatrix}, \qquad (D^2\Xi)_{(x,y)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad A(x, y) = (x, 2y),$$

$$(DA)_{(x,y)} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \qquad T_{(x,y)} = \begin{cases} \langle (1,0)\rangle \cup \langle (0,1)\rangle, & \text{if } (x, y) = (0, 0) \\ \langle (x, -y)\rangle, & \text{otherwise.} \end{cases}$$

It is now easy to check that condition (7) is satisfied as

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}[(x, 2y), \boldsymbol{w}] + \begin{bmatrix} y & x \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \cdot \boldsymbol{w} = 2yw_1 + xw_2 + yw_1 + 2xw_2 = 3(yw_1 + xw_2)$$

equals zero for all $\boldsymbol{w} \in T_{(x,y)}$. In fact, both the initial and the coarse grained system can be explicitly solved, the solutions being $\Phi_t(x, y) = (xe^t, ye^{2t})$ and $\Psi_t(z) = ze^{3t}$. It is now apparent that $\Xi \circ \Phi_t = \Psi_t \circ \Xi$, for all $t \geq 0$.

Another way to prove Theorem 6 is to consider discrete systems derived from the given continuous one and then require that all of those have the same coarse graining. The statement of the theorem will then follow directly from Theorem 4.

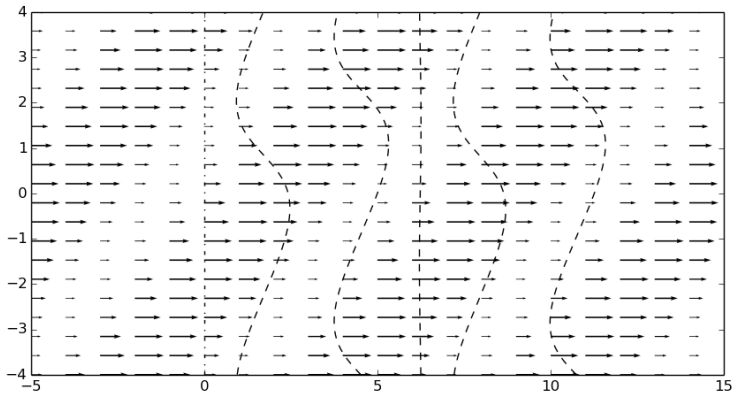

FIGURE 1. The dashed curves depict the evolution of the dotted line at times $\tau/4$, $3\tau/4$, τ, $5\tau/4$, and $7\tau/4$

We give a short sketch of the idea. Assume (X, Φ) is some smooth system. Fix any $\tau > 0$ and let (X, Ψ_1) be a discrete system generated with the map $\Psi_1 = \Phi_\tau$. No matter which τ was chosen, any coarse graining of Φ will coarse grain Ψ_1 as well. And conversely, if $\Xi\colon X \to Y$ coarse grains the system Φ_τ regardless of which $\tau > 0$ was chosen, then it coarse grains the initial continuous system (X, Ψ) by definition. Rewriting the condition from Theorem 4 in the form

$$(D\Xi)_{\Phi_{\tau(\boldsymbol{x})}} \cdot (D\Phi_\tau)_{\boldsymbol{x}} \cdot \boldsymbol{w} = 0, \text{ for all } \boldsymbol{w} \in T_{\boldsymbol{x}}$$

and then differentiating it with respect to τ and finally letting $\tau \to 0$ gives the condition from Theorem 6 as required.

However, if Ξ coarse grains only some of the induced discrete system then it need not coarse grain the continuous system as the following example shows.

Example 3. Consider a continuous system in the plane given by the following differential equation

$$\dot{x} = 2 + \sin(x - y), \qquad \dot{y} = 0.$$

It is possible to obtain the explicit solution for $\Phi_t(x, y)$ but this will not be necessary to make our point. It suffices to notice that all the trajectories of this system run parallel to the x-axis and that the vector field of their velocities, represented by the equation above, is periodic in the first variable with period 2π. Moreover, it is constant along the lines enclosing an angle of $\pi/4$ with the x-axis (see Figure 1).

Let us denote by τ the time taken by a particle carried by this vector field to travel from $(0, 0)$ to $(2\pi, 0)$, i.e. let $\tau > 0$ be such that $\Phi_\tau(0, 0) = (2\pi, 0)$. Because of the symmetries described above, it is now apparent that in fact $\Phi_\tau(x, y) = (2\pi + x, y)$ holds for any $(x, y) \in \mathbb{R}^2$. In particular, this discretisation is coarse grained by the map $\Xi(x, y) = x$. Nevertheless, this Ξ does not coarse grain the continuous system

as one can see using Theorem 6. We compute

$$(D\Xi)_{(x,y)} = \begin{bmatrix} 1 & 0 \end{bmatrix},$$
$$(D^2\Xi)_{(x,y)} = \mathbf{0},$$
$$(DA)_{(x,y)} = \begin{bmatrix} \cos(x-y) & -\cos(x-y) \\ 0 & 0 \end{bmatrix},$$
$$T_{(x,y)} = \langle (0,1) \rangle,$$

and substituting everything into the left hand side of (7) we get

$$-\cos(x-y)$$

which is clearly not equal to zero in the whole plane.

3. Motivating examples

In this section we describe a class of systems to which we wish to apply the theory of coarse grainings. In the theory of genetic algorithms they are better known as *binary tournaments*. In some special cases their coarse grainings have already been investigated in [16, 17].

Adopting the notation from [20], let $\Omega = \{1, 2 \ldots, n\}$ be a set of labels. Imagine a well mixed solution of a large number of particles in a tank each carrying a label from Ω. They move freely about the tank and every time two of them bump into each other their states are changed according to some rule. A rule, for example, might be that the particle in the lower state gets excited to match the state of the particle in the higher state (see Example 4 below).

If we denote the family of two point subsets of Ω with $[\Omega]^2$, then any function $s\colon [\Omega]^2 \to \Omega$ such that $s(S) \in S$ is called a *two point selection*. Equivalently s can be interpreted as an assignment of a direction to every edge of a complete graph with the set of vertices Ω. Intuitively, s is a rule to decide a winner among any pair of agents each of which is in a state from Ω. The fact $s(\{i, j\}) = i$, that is i beats j, will be also denoted as $i \leftharpoondown j$ in agreement with the directed graph interpretation, and to match the fact that the losing particle's label is transformed from j to i.

Associated to the selection s is a map $T\colon \Lambda_n \to \Lambda_n$ given by

$$(T(p_1, \ldots, p_n))_i = p_i(p_i + 2\sum_{i \leftharpoondown k} p_k), \tag{10}$$

which is often called a *heuristic*. The interpretation is as follows; $(T(\boldsymbol{p}))_i$ gives the probability that, after choosing at random two agents from the population in which the proportions of labels are given by $\boldsymbol{p} = (p_1, \ldots, p_n)$, the winner is of type i. In other words, the heuristic map T models (in discrete steps) the evolution of the particles in the tank under the assumption that the solution is well mixed and when the number of particles tends to infinity. This construction already generalises many of the examples of binary tournaments given in [17].

We wish to investigate coarse grainings of the map T. In particular, we will be interested in a special class of linear coarse grainings arising from partitions of the set Ω called *aggregations* or *lumpings*. The reason for this, as the following sections will show, is twofold. Firstly, for these we can provide a nice characterisation in terms of the parameters of the model, and secondly they automatically induce coarse grainings for a wide class of different systems that represent different ways to model (or simulate) the evolution of the solution in the tank.

Definition 7. Let $\mathcal{P} = \{C_1, \dots, C_m\}$ be a partition of the set $\Omega = \{1, 2, \dots, n\}$. The *aggregation* associated with this partition is a map $\Xi\colon \Lambda_n \to \Lambda_m$ defined by

$$\Xi \cdot (p_1, \dots, p_n)^T = (\sum_{i \in C_1} p_i, \dots, \sum_{i \in C_m} p_i)^T. \tag{11}$$

Hence, Ξ is a 0-1 matrix having exactly one entry equal to 1 in each column, i.e. it is stochastic. Given a heuristic $T\colon \Lambda_n \to \Lambda_n$, the partition $\mathcal{P}$ (or the induced equivalence relation) is said to be *compatible (with the dynamics T)* if Ξ is a coarse graining of the map T.

Definition 8. The equivalence relation $\equiv$ on the set of labels Ω is *contiguous* with respect to a selection $\leftharpoonup$ if for all $i, j, k \in \Omega$ we have

$$i \equiv j \equiv k \quad \text{whenever} \quad i \equiv k \text{ and } i \leftharpoonup j \leftharpoonup k. \tag{12}$$

Note that i and k in the definition above are distinct as we cannot have $i \leftharpoonup j$ and at the same time $j \leftharpoonup i$. Also, in general, the relation $\leftharpoonup$ need not be transitive. An example of a non-transitive rule is the game 'Rock, paper, scissors' and extensions thereof. We can now state our characterisation of compatible aggregations for these systems. Its proof is omitted as it follows from more general Proposition 11 we prove below.

Proposition 9. *Let T be a heuristic as above. The equivalence relation on Ω is compatible (i.e. gives a coarse graining) with T if and only if it is contiguous with respect to $\leftharpoonup$.*

Example 4 (Simple binary tournament, [17])**.** Let the relation $\leftharpoonup$ on $\{1, 2 \dots, n\}$ be given by

$$i \leftharpoonup j \quad \text{if and only if} \quad i > j$$

i.e. the particle in the lower state gets excited to match the state of the particle in the higher state. It is not hard to identify all the contiguous partitions for this relation. For any two $i, j \in \Omega$ that are in the same class, any k that lies in between $i > j > k$ must belong to the same class. Hence, the only contiguous partitions are those that can be obtained by writing numbers $1, 2, \dots, n$ in the increasing order and then making a few cuts to yield blocks of a partition. Proposition 9 says that these are the only compatible aggregations for the heuristic T given by

$$(T(p_1, \dots, p_n))_i = p_i(p_i + 2\sum_{i \leftharpoonup k} p_k) = p_i^2 + 2\sum_{k=1}^{i-1} p_i p_k, \qquad \text{for } 1 \le i \le n.$$

Suppose that it is not a selection that is given on Ω but let $P\colon \Omega^2 \to [0, 1]$ be a function such that $P(i, j) = 1 - P(j, i)$ for all $i, j \in \Omega$. We call such a P a *selection map*. The interpretation is that in a clash between i and j, i wins with probability $P(i, j)$ and j wins with the probability $1 - P(i, j) = P(j, i)$. Note that this requires $P(i, i) = 1/2$ for all $i \in \Omega$, the interpretation being that between two particles labelled by i, each wins half of the time. The outcome is, of course, the same either way. If $P(i, j) \in \{0, 1\}$ whenever $i \neq j$ then this reduces to a two point selection given by $s(\{i, j\}) = iP(i, j) + jP(j, i)$. It turns out that there is a characterisation of admissible aggregations that is the same as the one given in Proposition 9. Of course, the definition of a contiguous partition requires some amendments.

Definition 10. The equivalence relation $\equiv$ on the set of labels Ω is *contiguous* with respect to a selection map P if for all $i, j, k \in \Omega$ we have

$$i \equiv j \equiv k \quad \text{whenever} \quad i \equiv k \text{ and } P(i,j) \neq P(k,j). \tag{13}$$

We remark that this reduces to Definition 8 when P maps off-diagonal pairs to $\{0,1\}$.

If we denote by $\mathcal{P}$ the partition induced by the classes of $\equiv$ then the condition above amounts to requiring that for any $C, D \in \mathcal{P}$, $C \neq D$, and for any $i, k \in C$ and $j \in D$ we have $P(i,j) = P(k,j)$. If we let l be any other element in D, the same condition now applied with the roles of C and D swapped gives $P(j,k) = P(l,k)$. But since P is a selection this implies $P(k,j) = P(k,l)$ and hence $P(i,j) = P(k,l)$. This shows that for a contiguous partition there is a well defined selection map $P\colon \mathcal{P}^2 \to [0,1]$ such that $P(C,D) + P(D,C) = 1$ for all $C, D \in \mathcal{P}$, and for every $i \in C$ and $j \in D$ we have $P(C,D) = P(i,j)$. Although these are two different maps, we will keep the same notation.

How does the heuristic map look in this more general setting? Let us compute the probability that the winner in a random reaction is of type i.

$$(T(p_1, \ldots, p_n))_i = p_i^2 + \sum_{k \in \Omega \setminus \{i\}} 2 p_i p_k P(i,k) = 2 p_i \sum_{k \in \Omega} P(i,k) p_k, \tag{14}$$

where we noted that $P(i,i) = 1/2$. It is not hard to see that when we have a two point selection this reduces to (10). We are now ready to prove our characterisation.

Proposition 11. *Let P be a selection map on Ω, and let the heuristic T be as in* (14)*. The equivalence relation on Ω is compatible (i.e. gives a coarse graining) with T if and only if it is contiguous with respect to P.*

Proof. Let $\mathcal{P}$ be the partition of Ω induced by a contiguous equivalence relation, and let Ξ be the associated aggregation. Take $C \in \mathcal{P}$ and calculate

$$((\Xi \circ T)(p_1, \ldots, p_n))_C = \sum_{i \in C} (T(p_1, \ldots, p_n))_i = \sum_{i \in C} 2 p_i \left(\sum_{k \in \Omega} P(i,k) p_k \right) =$$

$$= 2 \sum_{i \in C} p_i \left(\sum_{D \in \mathcal{P}} \sum_{k \in D} P(i,k) p_k \right) = 2 \sum_{i \in C} p_i \left(\sum_{D \in \mathcal{P}} P(C,D) \left(\sum_{k \in D} p_k \right) \right) =$$

$$= 2 \left(\sum_{i \in C} p_i \right) \left(\sum_{D \in \mathcal{P}} P(C,D) \left(\sum_{k \in D} p_k \right) \right) = 2 p_C \sum_{D \in \mathcal{P}} P(C,D) p_D = \tilde{T}(\Xi(p_1, \ldots, p_n))$$

where $p_C = \sum_{i \in C} p_i$ for any $C \in \mathcal{P}$. This shows that Ξ is a coarse graining of the system and, moreover, we see that the coarse grained map $\tilde{T}$ is in the same form as the original one. It expresses the rule of transformation for meta-particles that are given by the classes of the partition $\mathcal{P}$.

Conversely, suppose that $\mathcal{P}$ is a partition of Ω whose associated aggregation Ξ coarse grains T. We need to prove that $\mathcal{P}$ (i.e. the equivalence relation that it induces) is contiguous with respect to P. Take $C, D \in \mathcal{P}$, $C \neq D$. Following the discussion after Definition 10, it suffices to prove that for any $i, k \in C$ and $j \in D$ we have $P(i,j) = P(k,j)$. For the sake of getting a contradiction assume that there are $i, k \in C$ and $j \in D$ such that $P(i,j) < P(k,j)$.

To simplify the notation we can, without loss of generality, assume that $i = 1$, $k = 2$, $j = 3$. Note that

$$\Xi(1/2, 0, 1/2, 0, \dots, 0) = \Xi(0, 1/2, 1/2, 0, \dots, 0)$$

and since Ξ is a coarse graining we must also have

$$((\Xi \circ T)(1/2, 0, 1/2, 0, \dots, 0))_C = ((\Xi \circ T)(0, 1/2, 1/2, 0, \dots, 0))_C$$

which, after using the definition

$$((\Xi \circ T)(p_1, \dots, p_n))_C = \sum_{i \in C} 2p_i \left(\sum_{k \in \Omega} P(i,k) p_k \right),$$

becomes

$$\frac{1}{2}(P(1,1) + P(1,3)) = \frac{1}{2}(P(2,2) + P(2,3)).$$

As $P(1,1) = P(2,2) = 1/2$, we get $P(1,3) = P(2,3)$, which contradicts the initial assumption $P(1,3) < P(2,3)$. This finishes the proof of the other implication. □

4. Random Heuristic Search

In this section we further generalise the model from Section 3 to include even more general reaction rules. But first, we fix some notation.

As before, we use a bold variable when we want to emphasise that the object is a vector, e.g. $\boldsymbol{p} = (p_1, \dots, p_n)^T \in \mathbb{R}^n$, and $\boldsymbol{e_i}$ is i-th vector of the canonical basis in $\mathbb{R}^n$. The set of non negative integers is denoted by $\mathbb{Z}_+$. Variables $\boldsymbol{v}$, $\boldsymbol{w}$, etc. are reserved for labelling vectors in $\mathbb{Z}_+^n$ and $|\boldsymbol{v}| := v_1 + v_2 + \cdots + v_n$. Following the multi index notation due to Schwartz, the factorial function naturally extends over integral vectors via $\boldsymbol{v}! = v_1! \cdots v_n!$ and so do integral powers $\boldsymbol{p}^{\boldsymbol{v}} := p_1^{v_1} p_2^{v_2} \dots p_n^{v_n}$. It is also convenient to introduce a shorthand $C(n,r) := \binom{r+n-1}{r}$, and for the space of (real) $m \times n$ matrices $M_{m \times n}$. Recall that the $(n-1)$-dimensional simplex is

$$\Lambda_n := \{(p_1, \dots, p_n)^T \in \mathbb{R}^n : |\boldsymbol{p}| = 1 \text{ and } p_i \geq 0 \text{ for all } 1 \leq i \leq n\} \subset \mathbb{R}^n. \tag{15}$$

Any continuous simplex self map $T \colon \Lambda_n \to \Lambda_n$ is called a *heuristic*. We also use vectors to denote partial derivatives as in $\partial_{\boldsymbol{v}} T(\boldsymbol{p}) = \frac{d^{|\boldsymbol{v}|}}{dp_1^{v_1} \dots dp_n^{v_n}} T(p_1, \dots, p_n)$. The following class of Markov chains was introduced in [20].

Definition 12. Let $T \colon \Lambda_n \to \Lambda_n$ be a heuristic. A random heuristic search (RHS) with a positive integer parameter r is a discrete-time Markov chain with a state space $\frac{1}{r} X_n^r \subset \Lambda_n$ where X_n^r denotes the set of all possible vectors in $\mathbb{Z}_+^n$ that add up to r (there are $C(n,r)$ many of them), and where the transition probabilities are given by

$$\mathbb{P}\left[\frac{1}{r}\boldsymbol{v} \to \frac{1}{r}\boldsymbol{w} \right] = \frac{r!}{\boldsymbol{w}!} \left(T\left(\frac{1}{r}\boldsymbol{v} \right) \right)^{\boldsymbol{w}}. \tag{16}$$

We will sometimes refer to this Markov chain as an *augmented system* or sometimes a *simulation of size* r. As with any Markov chain, this augmented system can also be represented as a deterministic linear map given by a square matrix $T_{(r)}$ of size $C(n,r)$. Note that in order to do this, we need to fix some ordering on the set $\frac{1}{r} X_n^r$.

We interpret this as follows. If the tank containing particles is of fixed size and we assume that the total number r of the particles inside does not change over time then X_n^r represents the space of all the possible populations that the tank can contain. We further scale this by $1/r$ in order to embed the state space into Λ_n.

The definition above implies that, starting from a population $\boldsymbol{v}$, the transition probabilities are given by a multinomial random variable with the expectation $T\left(\frac{1}{r}\boldsymbol{v}\right)$ and the parameter r. In other words, if we start with the population $\boldsymbol{v}$, in order to obtain the population in the next time step we first compute $T\left(\frac{1}{r}\boldsymbol{v}\right)$, this represents the proportions of different labels in an urn from which we draw r of them independently with replacement to form the new generation $\boldsymbol{w}$.

A few remarks are in order. First note that the examples from Section 3 fit within this framework as it is not hard to see that the heuristics there map Λ_n into itself. Secondly, when the map T is given by some reaction rules, as in that section, then it is possible to give more a natural interpretation of (16). Starting with the population $\boldsymbol{v}$, select two particles i, j from it at random, look up the reaction rule for those two, and select one of them, as before, based on the probability $P(i,j) = 1 - P(j,i)$. Repeat this procedure r times and collate the products to form the next generation $\boldsymbol{w}$. Special heuristic maps of this sort fall within the category of genetic algorithm operators, which have been, and still are, used to tackle optimisation problems by employing ideas from genetics such as selection, mutation and crossover. Each run of a particular genetic algorithm is simply a realisation of one trajectory of the Markov chain from the definition above. For a detailed account on this theory we refer the reader to [20].

There are two dynamic systems naturally arising in this discussion. The fist one is the discrete model generated by T with trajectories $p, T(p), T^2(p), \dots$ The other is a stochastic run of the induced Markov chain. Iterates of this RHS are sometimes denoted by $p, \tau(p), \tau^2(p), \dots$ but unlike the deterministic iterates of T, these are random variables. A different way to write Definition 12 in a more compact way is to inductively define iterates of the augmented system of size r by

$$\begin{aligned} \tau^0(p) &= p \\ \tau^k(p) &= \frac{1}{r} Z^k_{T(\tau^{k-1}(p))}, \text{ for } k \in \mathbb{N} \end{aligned} \tag{17}$$

where $\{Z^k_{\boldsymbol{\alpha}} : k \in \mathbb{N}, \boldsymbol{\alpha} \in T(\frac{1}{r}X_n^r)\}$ is a countable family of independent multinomial variables with parameters r and $\boldsymbol{\alpha}$. The expectation of the sample $\tau(p)$ is therefore exactly $T(p)$ but the actual distribution will depend on the size r.

It can be shown that for any fixed number N, the first N iterates $\tau(p), \dots, \tau^N(p)$ of the RHS converge in probability to the iterates of T as $r \to \infty$. This is because the variability of the RHS (measured as the variance of the multinomial variable) decreases as r increases (Theorems 3.5 and 13.2 in [20]). However, it is not possible to achieve the convergence result for the whole trajectory and Example 5 below shows what can go wrong.

Example 5. Let $T\colon \Lambda_3 \to \Lambda_3$ be a constant map $T(p) = (1/3, 1/3, 1/3)^T =: \alpha$. For any $r \in \mathbb{N}$ and any $p \in \Lambda_3$, $\tau(p)$[4] is distributed over $\frac{1}{r}X_3^r$ as a (scaled) trinomial variable with the parameters r and α. As a matter of fact, by (17), any iterate $\tau^k(p)$ has the same distribution and furthermore they form a family of independent random variables. As a consequence of the weak law of large numbers,

[4]Note the implicit dependence of $\tau(p)$ on r.

$(\tau(p),\dots,\tau^N(p))$ converge in probability, as they should, to $(\alpha,\dots,\alpha)$ for any given $N\in\mathbb{N}$ as $r\to\infty$. However, if we consider the tail event $\{\limsup_k |\tau^k(p)-T^k(p)|>\epsilon\}=\{\limsup_k|\tau^k(p)-\alpha|>\epsilon\}$ we can easily compute that it occurs with probability 1 no matter which $r\in\mathbb{N}$ or $0<\epsilon<1/3$ we chose. This shows that the random sequence $(\tau(p),\tau^2(p),\dots)$ as a whole does not converge in probability to the sequence $(\alpha,\alpha,\dots)$ as $r\to\infty$, even though the convergence holds for the corresponding initial parts.

We will now see what can be said about coarse grainings of such systems. Recall that we started with a map T on Λ_n and produced a Markov chain $T_{(r)}$ on a significantly larger space $\Lambda_{C(n,r)}$. As one would expect, the coarse grainings of these two systems are related. In particular, we will again be interested in those coarse grainings arising from compatible aggregations.

As before an aggregation associated to a partition $\{C_1,\dots,C_m\}$ of $\{1,2,\dots,n\}$ is the map

$$\Xi\cdot(p_1,\dots,p_n)^T=(\sum_{i\in C_1}p_i,\dots,\sum_{i\in C_m}p_i)^T, \tag{18}$$

and it is said to be compatible with T if it is a coarse graining for T. For any positive integer r we also define the induced aggregation of $\frac{1}{r}X_n^r$, still denoted by Ξ, as a map $\Lambda_{C(n,r)}\to\Lambda_{C(m,r)}$ that is associated to the partition arising from an equivalence relation on the set $\frac{1}{r}X_n^r$ given by

$$\frac{\boldsymbol{v}}{r}\equiv\frac{\boldsymbol{w}}{r}\qquad\text{if and only if}\qquad\Xi\left(\frac{\boldsymbol{v}}{r}\right)=\Xi\left(\frac{\boldsymbol{w}}{r}\right). \tag{19}$$

In other words, two populations of size r are considered to be equivalent if they have the matching counts of the labels from each of the classes $C_1,\dots C_m$. One can think of C_i's as meta-labels and then $\Xi\colon\Lambda_n\to\Lambda_m$ simply maps distributions over the labels $\{1,2,\dots,n\}$ to those over the meta-labels $\{C_1,\dots C_m\}$. The overloaded $\Xi\colon\Lambda_{C(n,r)}\to\Lambda_{C(m,r)}$ extends this to the distributions over the populations of size r. The following result from [20] establishes the connection between coarse grainings of T and the augmented systems.

Theorem 13 ([20], Theorem 17.7)**.** *An aggregation Ξ is a coarse graining for a heuristic T if and only if for any chosen size of population r the induced aggregation Ξ coarse grains the associated augmented system.*

For the converse to hold it is necessary that the induced Ξ's are compatible with τ for any population size $r\in\mathbb{N}$. The following example shows that this condition cannot be weakened.

Example 6. Consider a system given by the heuristic

$$T(p_1,p_2,p_3)=(p_1^2,2p_1p_2+p_2^2,2p_1p_3+2p_2p_3+p_3^2). \tag{20}$$

This is in fact a binary tournament on the state space $\Omega=\{1,2,3\}$ given by the transitive two point selection $3\leftarrow 2\leftarrow 1$ in which the higher state always wins over the lower state. If we take the population size to be $r=1$ then the only particle in the tank can carry the label either 1, 2 or 3. Each of these three configurations is a fixed point of τ, and the transition matrix is the identity matrix. Therefore any partition, say $\{\{1,3\},\{2\}\}$, gives a compatible aggregation of the simulated system. This particular partition, however, is not compatible with T as can be seen using

Proposition 9. And it will not be compatible with augmented systems either for $r > 1$.

The assumption that Ξ is an aggregation and not just any coarse graining is also crucial for Theorem 13 to hold.

Example 7. Consider the heuristic $T(p_1, p_2, p_3) = (p_1, p_3, p_2)$. The projection map $\Xi(p_1, p_3, p_2) = p_2 - p_3$ is clearly a coarse graining albeit not induced by an equivalence relation. The associated simulation of size 2 is a Markov chain on the state space $\frac{1}{2}X_3^2$, where

$$X_3^2 = \{(2,0,0), (0,2,0), (0,0,2), (1,1,0), (1,0,1), (0,1,1)\},$$

and with the transition matrix[5]

$$\begin{bmatrix} 1 & \cdot & \cdot & 1/4 & 1/4 & \cdot \\ \cdot & \cdot & 1 & \cdot & 1/4 & 1/4 \\ \cdot & 1 & \cdot & 1/4 & \cdot & 1/4 \\ \cdot & \cdot & \cdot & \cdot & 1/2 & \cdot \\ \cdot & \cdot & \cdot & 1/2 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1/2 \end{bmatrix}$$

which is not compatible with the induced aggregation that merges the states $(0,1,1)$ and $(2,0,0)$.

At the end of this section we recall the vector version of the well-known multinomial theorem. We will need it later on and it is extremely easy to state using the introduced notation. The proof is left to the reader.

Proposition 14 (The vector multinomial theorem)**.** *Let $\Xi \in M_{m\times n}$ be a stochastic 0-1 matrix ($m \leq n$), and let $\boldsymbol{u} \in \mathbb{Z}_+^m$. For any $\boldsymbol{p} \in \mathbb{R}^n$ we have*

$$(\Xi \boldsymbol{p})^{\boldsymbol{u}} = \sum_{\boldsymbol{y}\in\mathbb{Z}_+^n \,:\, \Xi(\boldsymbol{y})=\boldsymbol{u}} \frac{\boldsymbol{u}!}{\boldsymbol{y}!}\boldsymbol{p}^{\boldsymbol{y}}. \tag{21}$$

5. Aggregation coarse grainings

We will now prove our main theorem, a characterisation of compatible aggregations for the class of analytic heuristics in terms of their parameters. But first, we need the following lemma.

Lemma 15. *Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be a smooth function of class C^d and let $\Xi\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. If Ξ is a coarse graining of the system T then it coarse grains $\partial_{\boldsymbol{v}}T\colon \mathbb{R}^n \to \mathbb{R}^n$ for any $\boldsymbol{v} \in \mathbb{Z}_+^n$ for which $|\boldsymbol{v}| \leq d$.*

Proof. It is enough to prove that $\frac{d}{dp_i}T$ is coarse grained via Ξ for any $1 \leq i \leq n$ as the general result will then follow by induction. We denote the coarse grained system by $\tilde{T}$ and calculate

$$\Xi\frac{d}{dp_i}T(\boldsymbol{p}) = \Xi \lim_{h\to 0}\frac{1}{h}\left(T(\boldsymbol{p}+h\boldsymbol{e_i}) - T(\boldsymbol{p})\right) = \lim_{h\to 0}\frac{1}{h}\left(\Xi T(\boldsymbol{p}+h\boldsymbol{e_i}) - \Xi T(\boldsymbol{p})\right) =$$
$$= \lim_{h\to 0}\frac{1}{h}\left(\tilde{T}(\Xi\boldsymbol{p}+h\Xi\boldsymbol{e_i}) - \tilde{T}(\Xi\boldsymbol{p})\right) = (D\tilde{T})_{\Xi\boldsymbol{p}}\Xi\boldsymbol{e_i}$$

which is a function of $\Xi\boldsymbol{p}$ as can be seen from the formula. □

[5]The dots stand for zeros.

Note that the linearity of both the derivative and Ξ is crucial here as it has allowed us to swap the operators $\partial_{\boldsymbol{e_i}}$ and Ξ. We also note in passing that the calculation above proves that a surjective linear coarse graining of a smooth map produces a system with the same degree of smoothness. In fact, as Tóth et al. prove in [19] (Theorem 2.1), the Lipschitz property is also preserved in the coarse grained system, and with the same order. We now proceed to prove the main result.

Theorem 16. *Let $T(\boldsymbol{p}) = \sum_{\boldsymbol{v}\in\mathbb{Z}_+^n} \frac{1}{\boldsymbol{v}!}\boldsymbol{\alpha}_{\boldsymbol{v}}\boldsymbol{p}^{\boldsymbol{v}}$ be an absolutely convergent series with the infinite radius of convergence defining an analytic function on $\mathbb{R}^n$. An aggregation $\Xi\colon \mathbb{R}^n \to \mathbb{R}^m$ is a valid coarse graining if and only if*

$$\Xi(\boldsymbol{\alpha}_{\boldsymbol{v}}) = \Xi(\boldsymbol{\alpha}_{\boldsymbol{w}}) \quad \text{whenever} \quad \boldsymbol{v},\boldsymbol{w} \in \mathbb{Z}_+^n \text{ such that } \Xi(\boldsymbol{v}) = \Xi(\boldsymbol{w}). \tag{22}$$

In particular, for all $\boldsymbol{u} \in \mathbb{Z}_+^m$ the coefficients $\boldsymbol{\beta}_{\boldsymbol{u}} = \Xi(\boldsymbol{\alpha}_{\boldsymbol{v}})$, where $\boldsymbol{v} \in \mathbb{Z}_+^n$ is chosen such that $\Xi(\boldsymbol{v}) = \boldsymbol{u}$, are well-defined. Moreover, the coarse grained system is a convergent series on $\mathbb{R}^m$ given by $Q(\boldsymbol{q}) = \sum_{\boldsymbol{u}\in\mathbb{Z}_+^m} \frac{1}{\boldsymbol{u}!}\boldsymbol{\beta}_{\boldsymbol{u}}\boldsymbol{q}^{\boldsymbol{u}}$.

Note the similarity between condition (22) above and (13). Indeed, it is not hard to see that the former generalises the latter when T is a heuristic induced by a two point selection, and hence the theorem above generalises Proposition 11.

Proof. Assume that $\Xi(\boldsymbol{v}) = \Xi(\boldsymbol{w})$ implies $\Xi(\boldsymbol{\alpha}_{\boldsymbol{v}}) = \Xi(\boldsymbol{\alpha}_{\boldsymbol{w}})$. We need to show that $\Xi(T(\boldsymbol{p}))$ is a function of $\Xi(\boldsymbol{p})$. We calculate

$$\Xi(T(\boldsymbol{p})) = \Xi\left(\sum_{\boldsymbol{v}\in\mathbb{Z}_+^n} \frac{1}{\boldsymbol{v}!}\boldsymbol{\alpha}_{\boldsymbol{v}}\boldsymbol{p}^{\boldsymbol{v}}\right) = \sum_{\boldsymbol{v}\in\mathbb{Z}_+^n} \frac{1}{\boldsymbol{v}!}\Xi(\boldsymbol{\alpha}_{\boldsymbol{v}})\boldsymbol{p}^{\boldsymbol{v}} = \sum_{\boldsymbol{u}\in\mathbb{Z}_+^m}\sum_{\boldsymbol{v}:\Xi(\boldsymbol{v})=\boldsymbol{u}} \frac{1}{\boldsymbol{v}!}\Xi(\boldsymbol{\alpha}_{\boldsymbol{v}})\boldsymbol{p}^{\boldsymbol{v}} =$$
$$= \sum_{\boldsymbol{u}\in\mathbb{Z}_+^m} \frac{1}{\boldsymbol{u}!}\boldsymbol{\beta}_{\boldsymbol{u}} \sum_{\boldsymbol{v}\colon \Xi(\boldsymbol{v})=\boldsymbol{u}} \frac{\boldsymbol{u}!}{\boldsymbol{v}!}\boldsymbol{p}^{\boldsymbol{v}} = \sum_{\boldsymbol{u}\in\mathbb{Z}_+^m} \frac{1}{\boldsymbol{u}!}\boldsymbol{\beta}_{\boldsymbol{u}}(\Xi\boldsymbol{p})^{\boldsymbol{u}} = Q(\Xi\boldsymbol{p}),$$

where we used the vector multinomial theorem.

Conversely, assume that the aggregation Ξ is a coarse graining of T. It suffices to show that $\Xi(\boldsymbol{v}) = \Xi(\boldsymbol{w})$ implies $\Xi(\boldsymbol{\alpha}_{\boldsymbol{v}}) = \Xi(\boldsymbol{\alpha}_{\boldsymbol{w}})$ when $\boldsymbol{v}-\boldsymbol{w} = \boldsymbol{e_i}-\boldsymbol{e_j}$, i.e. when the vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ differ by 1 at exactly two coordinates. For any other choice of an equivalent pair $\boldsymbol{v}$ and $\boldsymbol{w}$ we can construct a chain of intermediate equivalent vectors transforming $\boldsymbol{v}$ to $\boldsymbol{w}$ by changing one coordinate at the time. Applying the base case to each link in the chain will produce a chain of equalities giving the general case. For example, if the first three labels/coordinates belong to the same class with respect to Ξ then

$$(3,0,0,1) \rightsquigarrow (2,1,0,1) \rightsquigarrow (1,2,0,1) \rightsquigarrow (0,2,1,1),$$

is a valid sequence of transformations.

Assume, therefore, that for some $\boldsymbol{y},\boldsymbol{v},\boldsymbol{w} \in \mathbb{Z}_+^n$ we have $\boldsymbol{y} = \boldsymbol{v} - \boldsymbol{e_i} = \boldsymbol{w} - \boldsymbol{e_j}$, and $\Xi(\boldsymbol{v}) = \Xi(\boldsymbol{w})$. Using Lemma 15 we get that Ξ coarse grains $\partial_{\boldsymbol{y}}T$. Note that $\boldsymbol{e_i}-\boldsymbol{e_j} = \boldsymbol{v}-\boldsymbol{w} \in \ker\Xi$, and the necessary condition for coarse graining (6) implies that

$$(D(\partial_{\boldsymbol{y}}T))_{(0,\dots,0)}(\boldsymbol{e_i}-\boldsymbol{e_j}) = \partial_{\boldsymbol{e_i}}\partial_{\boldsymbol{y}}T(0,\dots,0) - \partial_{\boldsymbol{e_j}}\partial_{\boldsymbol{y}}T(0,\dots,0) =$$
$$= \partial_{\boldsymbol{v}}T(0,\dots,0) - \partial_{\boldsymbol{w}}T(0,\dots,0) \in \ker\Xi.$$

Hence,

$$\Xi(\partial_{\boldsymbol{v}}T(0,\dots,0)) = \Xi(\partial_{\boldsymbol{w}}T(0,\dots,0)).$$

But the series for T is given as a Taylor series about $(0,\dots,0)$ and therefore $\partial_{\boldsymbol{v}} T(0,\dots,0) = \boldsymbol{\alpha}_{\boldsymbol{v}}$ for all $\boldsymbol{v} \in \mathbb{Z}_+^n$. Thus $\Xi(\boldsymbol{\alpha}_{\boldsymbol{v}}) = \Xi(\boldsymbol{\alpha}_{\boldsymbol{w}})$, as required. □

Observation. Note that the proof works even if the series expansion for T is given about a point other than 0. This allows to extend the result to any analytic function T just by requiring that the condition (22) holds locally for a series expansion of T about any point in the domain.

Corollary 17. *Let $T(\boldsymbol{p}) = \sum_{\boldsymbol{v}\in\mathbb{Z}_+^n} \frac{1}{\boldsymbol{v}!}\boldsymbol{\alpha}_{\boldsymbol{v}}\boldsymbol{p}^{\boldsymbol{v}}$ be a Taylor series of an analytic function as in Theorem 16. An aggregation $\Xi\colon \mathbb{R}^n \to \mathbb{R}^m$ is a valid coarse graining if and only if it coarse grains each of the homogeneous components of T which are defined for each $d \in \mathbb{Z}_+$ by*

$$T_d(\boldsymbol{p}) = \sum_{\boldsymbol{v}\in\mathbb{Z}_+^n : |\boldsymbol{v}|=d} \frac{1}{\boldsymbol{v}!}\boldsymbol{\alpha}_{\boldsymbol{v}}\boldsymbol{p}^{\boldsymbol{v}}. \tag{23}$$

Proof. It suffices to note that $|\boldsymbol{v}| = |\Xi(\boldsymbol{v})|$ and similarly for $\boldsymbol{w}$, and hence $|\boldsymbol{v}| = |\boldsymbol{w}|$ whenever $\Xi(\boldsymbol{v}) = \Xi(\boldsymbol{w})$. Applying Theorem 16 twice will now give the result. □

As an application of Theorem 16, we characterise aggregations in three systems of artificial chemistry.

Example 8 (Chemical reactions of degree d)**.** Let each number in $\Omega = \{1, 2, \dots, n\}$ represent a different chemical. Assume that for each choice $\{c_1, c_2, \dots, c_d\}$ of d elements in Ω, where repetitions are allowed but the order of the elements is irrelevant, we are given a distribution $\boldsymbol{\tau}_{\boldsymbol{v}} = (\tau^1_{\boldsymbol{v}}, \tau^2_{\boldsymbol{v}}, \dots, \tau^n_{\boldsymbol{v}})$ over Ω whose entries are interpreted as the proportions of each of the chemicals produced by a chemical reaction involving the particles $c_1, \dots, c_d$ as reactants. We write this as

$$c_1 + c_2 \cdots + c_d \to \tau^1_{\boldsymbol{v}} \cdot 1 + \tau^2_{\boldsymbol{v}} \cdot 2 + \cdots + \tau^n_{\boldsymbol{v}} \cdot n \tag{24}$$

and the index $\boldsymbol{v} \in \mathbb{Z}_+^n$ is the vector containing the counts of different types of the reactants involved. Hence $|\boldsymbol{v}| = d$, and for any such $\boldsymbol{v} \in \mathbb{Z}_+^n$ the products of the corresponding reaction are given as a distribution $\boldsymbol{\tau}_{\boldsymbol{v}}$ over Ω.

We consider the following simulation. A reaction tank contains a large amount of particles each of which is a chemical from Ω. The proportion of each chemical in the tank is given by the probability vector $\boldsymbol{p} \in \Lambda_n$. We choose d particles from the container at random, find the appropriate reaction (denoted by $\boldsymbol{v}$), and then produce a new particle $i \in \Omega$ according to the probabilities given by $\boldsymbol{\tau}_{\boldsymbol{v}}$. We repeat this procedure many times and collect all the products to form a new generation of particles in the tank.

What we have just described is precisely a RHS where the heuristic function is given by

$$T(\boldsymbol{p}) = \sum_{\boldsymbol{v}\in\mathbb{Z}_+^n : |\boldsymbol{v}|=d} \frac{1}{\boldsymbol{v}!}\left(d! \cdot \boldsymbol{\tau}_{\boldsymbol{v}}\right)\boldsymbol{p}^{\boldsymbol{v}}. \tag{25}$$

The factor $\frac{d!}{\boldsymbol{v}!}$ above accounts for the number of different ways to choose the reactants represented by $\boldsymbol{v}$, given that we are choosing d particles from the tank independently, with replacement, and with repetitions allowed. If we now set $\boldsymbol{\alpha}_{\boldsymbol{v}} = d! \cdot \boldsymbol{\tau}_{\boldsymbol{v}}$, we can directly apply Theorem 16 and conclude that in order for Ξ to be a compatible aggregation it must be $\Xi(\boldsymbol{\tau}_{\boldsymbol{v}}) = \Xi(\boldsymbol{\tau}_{\boldsymbol{w}})$ whenever $\Xi(\boldsymbol{v}) = \Xi(\boldsymbol{w})$. In other

words, a partition of Ω is compatible with the procedure described above, if taking the blocks of the partition as meta-chemicals, the distributions over these are well-defined for all the chemical reactions involving a multiset of d meta-chemicals as reactants.

Example 9. (Discrete-time homogeneous Markov chains) If we set $d = 1$ in the previous example, then for each label $i \in \Omega$ the vector $\boldsymbol{\tau}_{\boldsymbol{e_i}}$ gives the probabilities of the reaction

$$i \to \tau^1_{\boldsymbol{e_i}} \cdot 1 + \tau^2_{\boldsymbol{e_i}} \cdot 2 + \cdots + \tau^n_{\boldsymbol{e_i}} \cdot n. \tag{26}$$

These can be interpreted as the transition probabilities of a Markov chain, and one can see that the heuristic of this system is a linear operator on $\mathbb{R}^n$ corresponding to the Markov transition matrix of the chain. The compatibility criterion for aggregations in this case becomes the well-known condition for lumping the states of a Markov chain (see e.g. [9]). Any two states that are lumped together must have exactly the same outgoing transition probabilities towards any of the blocks in the partition.

Recall the system in Example 4. The heuristic there can be interpreted as a second degree chemistry where each pair of chemicals reacts to produce a chemical with the higher of the two reacting labels. We can, therefore, use the previous example to confirm our findings. Assume that $\Xi\colon \mathbb{R}^n \to \mathbb{R}^m$ is a compatible aggregation of this heuristic. By Example 8 the reactions among the meta-particles must be well-defined. This means that if two elements i and j (say $i < j$) belong to the same class, then any element $i < k < j$ in between would also have to be in the same class. Otherwise, the meta-reaction cannot be defined as $k + i \to 1 \cdot k$ but $k + j \to 1 \cdot j$. Conversely, it is clear that any such contiguous partition is indeed a compatible aggregation.

Similarly, we could interpret any system given by a selection map as in (14), as a chemistry of degree 2. But the chemistries are still more general since they also incorporate systems as in the following example.

Example 10. Consider a second order chemistry on $\mathbb{Z}_n = \{0, 1, 2, \dots n-1\}$ with the reaction rules

$$i + j \to 1 \cdot (i +_n j), \qquad \text{for all } i, j \in \mathbb{Z}_n.$$

Following Example 8, a partition of $\mathbb{Z}_n$ will be a valid aggregation if and only for any two blocks, it is well defined which block they produce. Let us fix one of the blocks and denote it by C. Then for any element $d \in \Omega$ the set $(C +_n d)$ must be contained within one block. From here it is not hard to see that all the blocks are translates (or cosets) of the set $\{0, l, 2l, \dots, (n-l)\}$, where l is some positive divisor of n. We have therefore established a correspondence between the compatible aggregations and the divisors of n. In fact, both the compatible aggregations and the set of divisors of n come equipped with a natural partial order making them into lattices. It is not hard to see that these lattices are isomorphic and the refinement relation in the former corresponds to the relation 'is divisible by' in the latter.

6. Concluding remarks

In the present paper we gave a criterion for an aggregation of variables to be compatible with a map that is given by its series expansion; and applied it to artificial

chemistries. We also clarified the distinction between discrete- and continuous-time systems when it comes to finding compatible aggregations.

Despite the attractive and deceptively simple looking characterisations presented here, the consideration of algorithms led the authors to the following significant conclusion. The existence of a non-trivial compatible aggregation for a Markov chain (a chemistry of degree 1) is an NP-complete problem. This result will appear in a forthcoming paper. This means that for some systems, an algorithm searching for compatible aggregations cannot essentially be better then the one that exhaustively tests all the partitions of the set of variables. But even if finding aggregations is computationally hard, the effort is usually worthwhile as the reduced system, once found, allows for significant speed-ups when computing individual trajectories of the system.

One way to overcome this problem is to look for compatible aggregations that refine a given non-trivial partition. This approach has previously been employed in model reduction algorithms for DTMCs, one of such being 'partition minimisation' introduced by Paige and Tarjan in [14], and adapted to Markov chains in [3].

Let us stress again that, given a heuristic function, RHS as defined above is just one possible way of simulating the system the heuristic is representing. The main characteristic of the RHS approach is that it is generational. Each new generation, although dependent on the previous one, consists of an entirely new collection of entities. Sometimes, other approaches may be more appropriate. Indeed, if our algorithm is supposed to simulate a real biological system, it is natural to allow the members of different generations to coexist. Similarly, if we are to simulate a chemical process, we would like to keep most of the particles for the new generation, and only those few that bump into another particle with which they can react, will be replaced with a product of the reaction in the next time step. One such natural model is given in Appendix A but it would be interesting to see which other approaches could work.

Artificial chemistries are a fruitful area and it is no surprise that different authors (Fontana et al. [5], Tóth et al. [19], Dittrich et al. [4] to name a few) model chemical reactions in different ways. What seems promising is that all of them incorporate a polynomial model similar to the one we had above. This should mean that the results on compatible aggregations can be easily transferred across the modelling paradigms, and this is something we would like to explore in future.

Acknowledgements

The authors acknowledge support from the European Union through funding under FP7-ICT-2011-8 project HIERATIC (316705).

References

[1] L. S. Block and W. A. Coppel, *Dynamics in one dimension*, Lecture Notes in Mathematics, vol. 1513, Springer-Verlag, Berlin, 1992. MR1176513 (93g:58091)

[2] Keki Burjorjee, *Sufficient conditions for coarse-graining evolutionary dynamics*, Foundations of genetic algorithms, 2007, pp. 35–53.

[3] S. Derisavi, H. Hermanns, and W. Sanders, *Optimal state-space lumping in Markov chains*, Information Processing Letters **87** (2003September), no. 6, 309–315.

[4] Peter Dittrich and Pietro Speroni Di Fenizio, *Chemical organisation theory*, Bulletin of mathematical biology **69** (2007), no. 4, 1199–1231.

[5] Walter Fontana and Leo W Buss, *The arrival of the fittest: Toward a theory of biological organization*, Bulletin of Mathematical Biology **56** (1994), no. 1, 1–64.

[6] Aleksandr Nikolaevich Gorban, Nikolas K Kazantzis, Ioannis G Kevrekidis, Hans Christian Öttinger, and Constantinos Theodoropoulos, *Model reduction and coarse-graining approaches for multiscale phenomena*, Springer, 2006.
[7] Martin N. Jacobi, *Hierarchical organization in smooth dynamical systems*, Artif. Life **11** (December 2005), no. 4, 493–512.
[8] Martin Nilsson Jacobi and Olof Görnerup, *A spectral method for aggregating variables in linear dynamical systems with application to cellular automata renormalization*, Advances in Complex Systems **12** (2009), no. 02, 131–155.
[9] John G. Kemeny and J. Laurie Snell, *Finite Markov chains*, Springer-Verlag, New York-Heidelberg, 1976. Reprinting of the 1960 original, Undergraduate Texts in Mathematics. MR0410929 (53 #14670)
[10] Genyuan Li and Herschel Rabitz, *A general analysis of exact lumping in chemical kinetics*, Chemical engineering science **44** (1989), no. 6, 1413–1430.
[11] ———, *A general analysis of approximate lumping in chemical kinetics*, Chemical engineering science **45** (1990), no. 4, 977–1002.
[12] Genyuan Li, Herschel Rabitz, and János Tóth, *A general analysis of exact nonlinear lumping in chemical kinetics*, Chemical Engineering Science **49** (1994), no. 3, 343–361.
[13] J. R. Norris, *Markov chains*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 2, Cambridge University Press, Cambridge, 1998. Reprint of 1997 original. MR1600720 (99c:60144)
[14] Robert Paige and Robert E. Tarjan, *Three partition refinement algorithms*, SIAM J. Comput. **16** (1987), no. 6, 973–989. MR917035 (89h:68069)
[15] Jonathan E. Rowe, Michael D. Vose, and Alden H. Wright, *Coarse graining selection and mutation*, Foundations of genetic algorithms, 2005, pp. 176–191. MR2238251
[16] ———, *State aggregation and population dynamics in linear systems*, Artif. Life **11** (December 2005), no. 4, 473–492.
[17] ———, *Differentiable coarse graining*, Theoret. Comput. Sci. **361** (2006), no. 1, 111–129. MR2254227 (2007h:68171)
[18] Hal L. Smith and Horst R. Thieme, *Dynamical systems and population persistence*, Graduate Studies in Mathematics, vol. 118, American Mathematical Society, Providence, RI, 2011. MR2731633 (2012c:92060)
[19] Alison S Tomlin, Genyuan Li, Herschel Rabitz, and János Tóth, *The effect of lumping and expanding on kinetic differential equations*, SIAM Journal on Applied Mathematics **57** (1997), no. 6, 1531–1556.
[20] Michael D. Vose, *The simple genetic algorithm*, Complex Adaptive Systems, MIT Press, Cambridge, MA, 1999. Foundations and theory, A Bradford Book. MR1713436 (2000h:65024)

Appendix A. Different models

In this section we look into a different way of modelling chemical reactions, and show that, when it comes to compatible aggregations, the same results we obtained before still hold. The idea is to start with an abstract model of reactions of degree d. This can be given by a set of rules as in (24) or concisely written as a heuristic (25). In Section 4 a procedure was given for any $r \in \mathbb{N}$ to construct a Markov chain on the state space $\frac{1}{r}X_n^r$ which was called a simulation or RHS of size r. We have there seen that the aggregations compatible with the heuristic extend naturally to aggregations of the associated RHS. The purpose of this section is to show that for another natural derived model this still holds.

The approach in Section 4 was generational. With each time step the whole generation would have been replaced. Consider now a model where in each time step only one interaction occurs between d agents chosen at random. As before $\Omega = \{1, 2, \dots, n\}$, and let

$$T(\boldsymbol{p}) = \sum_{\boldsymbol{v} \in X_n^d} \frac{1}{\boldsymbol{v}!} \left(d! \cdot \boldsymbol{\tau}_{\boldsymbol{v}}\right) \boldsymbol{p}^{\boldsymbol{v}}$$

be a heuristic, and let $r \in \mathbb{N}$. We wish to construct a Markov chain with the set of states $\frac{1}{r}X_n^r$ so that the transitions model the following process. Starting with a state $\frac{1}{r}\boldsymbol{v}$ representing the initial population $\boldsymbol{v} \in X_n^r$, we firstly choose d elements at random from $\boldsymbol{v}$ (repetitions are allowed). Then we look up the reaction involving those reactants $\boldsymbol{u} \in X_n^d$ and one product is chosen according to the weights given by $\boldsymbol{\tau_u}$. Finally, one element from $\boldsymbol{v}$ is chosen at random to be replaced with the produced particle. This population, after scaling by $\frac{1}{r}$, represents the state in the following time step.

It is straightforward to compute the transitions for this model,

$$\mathbb{P}\left[\frac{1}{r}\boldsymbol{v} \to \frac{1}{r}\boldsymbol{w}\right] = \begin{cases} \sum_{i=1}^{n} \left(T\left(\frac{1}{r}\boldsymbol{v}\right)\right)_i \frac{1}{r}v_i, & \text{if } \boldsymbol{w} = \boldsymbol{v} \\ \left(T\left(\frac{1}{r}\boldsymbol{v}\right)\right)_j \frac{1}{r}v_i, & \text{if } \boldsymbol{w} - \boldsymbol{v} = \boldsymbol{e}_j - \boldsymbol{e}_i \text{ and } i \neq j \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

However, it will be convenient to write this Markov chain in a different way. Take a countable family of independent random variables

$$\left\{ Z_{\boldsymbol{\alpha}}^k, Y_{\boldsymbol{\beta}}^k : k \in \mathbb{N}, \boldsymbol{\alpha} \in T\left(\frac{1}{r}X_n^r\right), \boldsymbol{\beta} \in \frac{1}{r}X_n^r \right\},$$

where

$$Z_{\boldsymbol{\alpha}}^k \sim \begin{pmatrix} 1 & 2 & \dots & n \\ \alpha_1 & \alpha_2 & \dots & \alpha_n \end{pmatrix} \text{ and } Y_{\boldsymbol{\beta}}^k \sim \begin{pmatrix} 1 & 2 & \dots & n \\ \beta_1 & \beta_2 & \dots & \beta_n \end{pmatrix}$$

and let $\frac{1}{r}V_0, \frac{1}{r}V_1, \frac{1}{r}V_2, \dots$ denote a sequence of random trajectories over $\frac{1}{r}X_n^r$ of this chain. Then it is clear that for all $k \in \mathbb{N}$ we have

$$\frac{1}{r}V_k = F\left(\frac{1}{r}V_{k-1}, Z_{T\left(\frac{1}{r}V_{k-1}\right)}^k, Y_{\frac{1}{r}V_{k-1}}^k\right), \tag{28}$$

where $F\colon \frac{1}{r}X_n^r \times \Omega \times \Omega \to \frac{1}{r}X_n^r$ is given by

$$F\left(\frac{1}{r}\boldsymbol{v}, i, j\right) = \frac{1}{r}\left(\boldsymbol{v} + \boldsymbol{e_i} - \boldsymbol{e_j}\right), \text{ for all } \boldsymbol{v} \in X_n^r, i, j \in \Omega. \tag{29}$$

Let $\pi\colon \Omega \to \{C_1, \dots, C_m\} = \tilde{\Omega}$ be a partition map and $\Xi\colon \Lambda_n \to \Lambda_m$ the associated aggregation. In particular $\Xi\left(\frac{1}{r}X_n^r\right) = \frac{1}{r}X_m^r$. Then from the definition of F it is clear that a map $\tilde{F}\colon \frac{1}{r}X_m^r \times \tilde{\Omega} \times \tilde{\Omega} \to \frac{1}{r}X_m^r$ is well-defined by

$$\tilde{F}\left(\Xi\left(\frac{1}{r}\boldsymbol{v}\right), \pi(i), \pi(j)\right) = \Xi\left(F\left(\frac{1}{r}\boldsymbol{v}, i, j\right)\right) \tag{30}$$

or in other words, $\Xi \times \pi \times \pi$ is compatible with F.

Now it is easy to prove that the partition of $\frac{1}{r}X_n^r$ induced by Ξ is also a compatible aggregation of this derived chain. Employing the usual criteria for matching outgoing probabilities, it suffices to see that for any three states $\boldsymbol{v}, \boldsymbol{u}, \boldsymbol{w} \in X_n^r$ the equality

$$\mathbb{P}\left[\Xi\left(\frac{1}{r}V_k\right) = \Xi\left(\frac{1}{r}\boldsymbol{w}\right) \mid V_{k-1} = \boldsymbol{v}\right] = \mathbb{P}\left[\Xi\left(\frac{1}{r}V_k\right) = \Xi\left(\frac{1}{r}\boldsymbol{w}\right) \mid V_{k-1} = \boldsymbol{u}\right]$$

holds as soon as $\Xi\left(\frac{1}{r}\boldsymbol{v}\right) = \Xi\left(\frac{1}{r}\boldsymbol{u}\right)$. The LHS of this expression is

$$\begin{aligned}
\mathbb{P}\left[\Xi\left(F\left(\frac{1}{r}\boldsymbol{v}, Z^k_{T\left(\frac{1}{r}\boldsymbol{v}\right)}, Y^k_{\frac{1}{r}\boldsymbol{v}}\right)\right) = \Xi\left(\frac{1}{r}\boldsymbol{w}\right) \mid V_{k-1} = \boldsymbol{v}\right] = \\
= \mathbb{P}\left[\Xi\left(F\left(\frac{1}{r}\boldsymbol{v}, Z^k_{T\left(\frac{1}{r}\boldsymbol{v}\right)}, Y^k_{\frac{1}{r}\boldsymbol{v}}\right)\right) = \Xi\left(\frac{1}{r}\boldsymbol{w}\right)\right] = \\
= \sum_{C,D\in\tilde{\Omega}} \sum_{i\in C} \sum_{j\in D} \chi_{\Xi\left(F\left(\frac{1}{r}\boldsymbol{v},i,j\right)\right)=\Xi\left(\frac{1}{r}\boldsymbol{w}\right)} \mathbb{P}\left[Z^k_{T\left(\frac{1}{r}\boldsymbol{v}\right)} = i, Y^k_{\frac{1}{r}\boldsymbol{v}} = j\right] = \\
= \sum_{C,D\in\tilde{\Omega}} \sum_{i\in C} \sum_{j\in D} \chi_{\tilde{F}\left(\Xi\left(\frac{1}{r}\boldsymbol{v}\right),C,D\right)=\Xi\left(\frac{1}{r}\boldsymbol{w}\right)} \mathbb{P}\left[Z^k_{T\left(\frac{1}{r}\boldsymbol{v}\right)} = i\right] \mathbb{P}\left[Y^k_{\frac{1}{r}\boldsymbol{v}} = j\right] = \\
= \sum_{C,D\in\tilde{\Omega}} \chi_{\tilde{F}\left(\Xi\left(\frac{1}{r}\boldsymbol{v}\right),C,D\right)=\Xi\left(\frac{1}{r}\boldsymbol{w}\right)} \mathbb{P}\left[Z^k_{T\left(\frac{1}{r}\boldsymbol{v}\right)} \in C\right] \mathbb{P}\left[Y^k_{\frac{1}{r}\boldsymbol{v}} \in D\right],
\end{aligned}$$

where $\chi_{(\cdot)}$ represents an indicator, 0-1 function. Finally, it remains to notice that the last expression above will not change if we substitute $\boldsymbol{u}$ instead of $\boldsymbol{v}$. This is beacause $\Xi\left(\frac{1}{r}\boldsymbol{v}\right) = \Xi\left(\frac{1}{r}\boldsymbol{u}\right)$, and hence also $\Xi\left(T\left(\frac{1}{r}\boldsymbol{v}\right)\right) = \Xi\left(T\left(\frac{1}{r}\boldsymbol{u}\right)\right)$. Now backtracking the same steps we get the RHS.

The advantage of this approach is that many different derived models (simulations of size r) form a single heuristic can be represented in this way. And for all of them we have just shown that the aggregations that work for the heuristic, also coarse grain those simulated Markov chains. This justifies our primary concern of coarse graining heuristics.

Example 11. Let $\Omega = \{1, 2, 3\}$, $r = 2$ and let the reactions be given by the 'higher number wins' rule, $i + j \to 1 \cdot \max\{i, j\}$. The transition matrix of the RHS is

$$\begin{bmatrix}
1 & \cdot & \cdot & 1/16 & 1/16 & \cdot \\
\cdot & 1 & \cdot & 9/16 & \cdot & 1/16 \\
\cdot & \cdot & 1 & \cdot & 9/16 & 9/16 \\
\cdot & \cdot & \cdot & 6/16 & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & 6/16 & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & 6/16
\end{bmatrix},$$

where the state space is $\frac{1}{2}X^2_3$ and,

$$X^2_3 = \{(2,0,0), (0,2,0), (0,0,2), (1,1,0), (1,0,1), (0,1,1)\}.$$

The other model we described in this section has the same state space but the transitions are different

$$\begin{bmatrix}
1 & \cdot & \cdot & 1/8 & 1/8 & \cdot \\
\cdot & 1 & \cdot & 3/8 & \cdot & 1/8 \\
\cdot & \cdot & 1 & \cdot & 3/8 & 3/8 \\
\cdot & \cdot & \cdot & 1/2 & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & 1/2 & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & 1/2
\end{bmatrix}.$$

Two derived models exhibit different behaviours but nonetheless, the same aggregations are compatible with both systems, those induced by contiguous partitions.