CHAPTER SEVEN

FORMAL DYNAMICS

We develop in greater generality the participator dynamics introduced in special cases in chapter six.

1. Some fundamentals

In this chapter we develop the basic formalism for participator dynamics. Properly speaking, this is a dynamical system on a set of observers, namely the set $\mathcal{B}$ of objects of perception for an environment supported by a reflexive framework $(X, Y, E, S, \pi_\bullet)$ (**5**–2.6). However, given certain special assumptions which we review below, we can view the dynamics as taking place on E rather than $\mathcal{B}$. We are interested in the case where the reflexive framework is a symmetric framework (X, Y, E, S, G, J, π) (**5**–5.1); when appropriate we will indicate the special form that our emerging results and definitions take in this case. For simplicity, when we consider a symmetric framework *we assume throughout this chapter that E is a principal homogeneous J-space,* although J need not be abelian.

We first define the concept of *action kernel* at this level of generality. Intuitively, as discussed in **6**–2, an action kernel describes how a participator, at a given moment of reference time, changes perspective in response to a channeling. This change depends, in part, on the perspective of the participator, so that an action kernel is actually a family of kernels, one for each point of E. In the case of a symmetric framework we use the group action to define the notion of a *symmetric action kernel.* Such a kernel is generated from a single kernel on J, giving a symmetric description of the perspective-change law. In chapter six we studied a simplified version of this symmetric case.

Definition 1.1. (i) An *action kernel* on the reflexive framework (X, Y, E, S, $\pi_\bullet$) is a family $\{Q_e\}_{e\in E}$ of kernels on E such that, for each e, Q_e is a markovian kernel, $Q_e\colon E\times\mathcal{E}\to[0,1]$, and satisfies $Q_e(e_1,\cdot)=Q_e(e_2,\cdot)$ if $\pi_e(e_1)=\pi_e(e_2)$.

(ii) A *symmetric action kernel* on the symmetric observer framework (X, Y, E, S, G, J, π) is an action kernel $\{Q_e\}_{e\in E}$ with the following property: there exists a markovian kernel $Q\colon J\times\mathcal{J}\to[0,1]$ such that $Q(j,\cdot)=Q(j',\cdot)$ if $\pi(j)=\pi(j')$, and each Q_e is deduced from Q by the formula $Q_e(e_1,\Delta)=Q(e_1e^{-1},\Delta e^{-1})$. We say that *the symmetric action kernel is generated by* Q*.*

Here, as usual, e_1e^{-1} denotes the element of J which sends e to e_1; it is unique since we assume E is principal homogeneous for J. Similarly, for $\Delta\in\mathcal{E}$, Δe^{-1} is the set of all elements of J which send e into Δ.

In general we simply use the notation Q for the entire action kernel, so that Q stands for the whole family $\{Q_e\}_{e\in E}$. In the special case of symmetric action kernels the symbol Q denotes both the action kernel itself (i.e., the family of kernels $\{Q_e\}_{e\in E}$) and the kernel which generates it. However these notions contain the same information, so this abuse of language will not cause any problem. In the general case, i.e., part (i) of Definition 1.1, there are no compatibility requirements within the family $\{Q_e\}_{e\in E}$. In practice, however, the families that we consider have various kinds of internal consistency, but it is inappropriate to incorporate these in the basic definition.

We can now use the notion of *participator* as in Definition **6**–2.6. A participator is a triple $(\xi,\{Q(n)\}_n,\{\eta(n)\}_n)$, where ξ is a probability measure on E, $\{Q(n)\}_n$ is a sequence of action kernels, and $\{\eta(n)\}_n$ is a sequence of families of interpretation kernels for the reflexive framework. In this chapter we suppress mention of $\eta(n)$; in particular we do not consider the crucial question of the role played by the $\eta(n)$ in some generalized notion of action kernel. (Thus the dynamics we develop here is actually a preparticipator dynamics (**6**–2.6).) However, our present formalism does permit a participator's action kernel to vary in time, and intuitively $\eta(n)$ may be responsible in part for this evolution.

A participator manifests at each instant of reference time as an observer in the framework. The manifestation is probabilistic, so that we can think of the participator as a time sequence of random variables taking values in the set of observers in the framework. Since we suppress the interpretation kernels

$\eta(n)$, only the perspectives of the observers vary. Therefore we can think of a participator A as a sequence of random variables, say $W_1, W_2, \ldots$ taking values in E. ξ is then the distribution of W_0. $Q_e(n)(e_1, \cdot)$ is the distribution of W_{n+1} given that the value of W_n was e and given that the participator's manifestion at time n channeled with an observer in the framework whose perspective is represented by $e_1 \in E$ (via the map Π defining the framework as in **5**–2.2). Therefore, the process $W_1, W_2, \ldots$ is not a Markov chain: the distribution of W_{n+1} depends not just on the value e of W_n, but also on $\pi(e_1)$.

However, if we consider collectively a "closed system" of participators (**6**–2.9) and if we assume some regularity to the distribution of channelings, then we can canonically associate certain Markov chains to the system. These Markov chains contain complete information about the extended semantics for each potential manifestation of the participators. These potential manifestations together constitute the set of distinguished objects of perception $\mathcal{B}_E$ for the environment (**4**–4.4, **6**–2.7).

In this chapter we introduce and study certain Markov chains canonically associated to discrete-time, participator dynamical systems. We make the following restrictive assumptions:

1. The interpretation kernels $\eta(n)$ of the various participators have no explicit role in the dynamical formalism, so that we can view the participators as being individuated only by their perspectives.
2. The participator dynamical systems are ***closed*** (**6**–2.9).
3. The ***independent action*** postulate holds (**6**–2.10).
4. The choice of channeling at each instant of reference time may be described by a "τ-distribution" (defined in section two of this chapter).

To begin, we choose an integer $k \geq 1$, and consider k participators

$$A_i = \{(\xi_i,\ \{Q_i(n)\}_n)\},\ i = 1, \ldots, k$$

on our framework. We assume that there is a discrete time, at each instant of which each participator manifests as some observer in the framework, and at each instant of which there is a channeling; this is the same as the ***reference time*** of **6**–2. In effect we are studying an extended semantics in which these observer manifestations of the participators are the distinguished objects of perception. We make the ***closed system assumption*** (**6**–2.9) that at each instant the distinguished part of the channeling involves only observers which are participator manifestations at that instant. In other words, at any time n the only channelings involving participator manifestations are channelings between the participator manifestations themselves. Thus we assume that at any instant n of a participator channeling sequence (**6**–2.8), the domain D_n of the distinguished part of the channeling at that instant is a subset of the set of participator manifestations at time n; hence, D_n contains at most k observers. We need a consistent way to refer to the possible channelings among these observers:

Notation 1.2. Denote by $\mathcal{I}(k)$ the set of all involutions on subsets of $\{1,\ldots,k\}$. Thus an element of $\mathcal{I}(k)$ is a pair consisting of a subset $D \subset \{1,\ldots,k\}$ and a function $\chi: D \to D$ such that $\chi^2 \equiv \chi \circ \chi = \mathrm{id}_D$. More generally, if V is any set we denote by $\mathcal{I}(V)$ the set of all involutions of subsets of V. If V' and V'' are disjoint subsets of V with $\chi' \in \mathcal{I}(V')$ and $\chi'' \in \mathcal{I}(V'')$, we denote by $\chi' \cup \chi''$ the element of $\mathcal{I}(V' \cup V'')$ described by χ' on V' and by χ'' on V''. For $\chi \in \mathcal{I}(k)$, $D(\chi)$ denotes, as indicated above, the domain of χ.

Once we have fixed the integer k and the participators $A_1,\ldots,A_k$, the element χ of $\mathcal{I}(k)$ refers to the channeling in which, for each $i \in D(\chi)$, the manifestations of A_i and $A_{\chi(i)}$ channel to each other and in which, for $j \notin D(\chi)$, the manifestation of A_j does not channel. Henceforth, for simplicity, we will say "A_i and $A_{\chi(i)}$ channel to each other at time n" rather than the correct but cumbersome "the manifestations of A_i and $A_{\chi(i)}$ at time n channel to each other." Similarly, we will simply say "A_i has perspective (or 'position') e_i at time n" rather than "the manifestation of A_i at time n has perspective which is $\Pi(e_i)$." Thus, to say "$\chi \in \mathcal{I}(k)$ is the channeling at time n" means that, for $i \in D(\chi)$, A_i and $A_{\chi(i)}$ channel at time n, and for $j \notin D(\chi)$, A_j does not channel at time n. This is the same as saying that in the participator channeling sequence $(D_n,\chi_n) = (D(\chi),\chi)$. Thus, the $\mathcal{I}(k)$ notation permits us to consider channelings in which some participators are inactive, some channel to themselves, and so on.

As a result of a channeling at a given time t_0, each active participator changes its perspective, i.e., its position in E, in a manner dictated by its action kernel at time t_0. According to our participator notation the action kernel of A_i at time t_0 is $Q_i(t_0)$. Suppose the channeling is represented by $\chi \in \mathcal{I}(k)$. Then if $i \in D(\chi)$, the position of A_i at t_0+1, i.e., at the next instant of reference time, is a random variable with distribution $Q_{ie_i}(t_0)(e_{\chi(i)},\cdot)$. As in part (i) of Definition 1.1 above, $Q_{ie_i}(t_0)$ denotes the markovian kernel that governs the perspective change of A_i from time t_0 to t_0+1 given that A_i

has perspective e_i at t_0. In action kernel notation this means the following: for $\Delta_i \in \mathcal{E}$, the probability that the perspective of A_i will be in Δ_i at time $t_0 + 1$ is $Q_{ie_i}(t_0)(e_{\chi(i)}, \Delta_i)$. According to the independent action postulate, given the perspectives of the participators at time t_0 and given the channeling χ, the k E-valued random variables describing the next perspectives of the k participators are *independent*. As a result we have the following proposition:

Proposition 1.3. Suppose we are given, at time t_0, k participators $A_1, \ldots, A_k$ with action kernels $Q_1, \ldots, Q_k$. Moreover suppose that, at time t_0, A_i is at e_i, for $i = 1, \ldots, k$. Let $\chi \in \mathcal{I}(k)$ be the channeling at t_0, so that A_i and $A_{\chi(i)}$ channel to each other for each $i \in D(\chi)$ and so that A_j is inert if $j \notin D(\chi)$. Let $\Delta_1, \ldots, \Delta_k \in \mathcal{E}$, and let $N_{t_0,\chi}(e_1, \ldots, e_k; \Delta_1 \times \ldots \times \Delta_k)$ denote the probability that, at time $t_0 + 1$, A_i will have perspective in Δ_i, for $i = 1, \ldots, k$. Then

$$N_{t_0,\chi}(e_1, \ldots, e_k; \Delta_1 \times \ldots \times \Delta_k) = \prod_{i \in D(\chi)} Q_{ie_i}(t_0)(e_{\chi(i)}, \Delta_i) \prod_{j \notin D(\chi)} 1_{\Delta_j}(e_j).$$

In the case of symmetric action kernels the formula is

$$\prod_{i \in D(\chi)} Q_i(t_0)(e_{\chi(i)} e_i^{-1}, \Delta_i e_i^{-1}) \prod_{j \notin D(\chi)} 1_{\Delta_j}(e_j).$$

Proof. Straightforward. The independent action postulate justifies the products in the formulas above. ∎

We frequently express the formulae of 1.3 in ***infinitesimal form,*** replacing each Δ_i with dy_i and each $1_{\Delta_j}(e_j)$ with $\epsilon_{e_j}(dy_j)$. (Recall that $\epsilon_e(dy)$ is Dirac measure concentrated at e.) The first formula becomes

$$N_{t_0,\chi}(e_1, \ldots, e_k; dy_1, \ldots, dy_k) = \prod_{i \in D(\chi)} Q_{ie_i}(t_0)(e_{\chi(i)}, dy_i) \prod_{j \notin D(\chi)} \epsilon_{e_j}(dy_j).$$

In the symmetric case this is

$$\prod_{i \in D(\chi)} Q_i(t_0)(e_{\chi(i)} e_i^{-1}, dy_i e_i^{-1}) \prod_{j \notin D(\chi)} \epsilon_{e_j}(dy_j).$$

2. The τ-distribution

k participators, interacting via channeling, change their manifestations probabilistically in a manner governed by their action kernels; intuitively the result is a stochastic process indexed by reference time, with state space $\mathcal{B}^k$ (or more precisely $\mathcal{B}_E^k \subset \mathcal{B}^k$). Since we focus only on the ***perspectives*** of the manifestations, we can view this process as having state space E^k. We take our k participators to be $A_1, \ldots, A_k$ as above. Suppose we assume, artificially, that the same channeling pattern, say $\chi \in \mathcal{I}(k)$, occurs at each instant of reference time. In other words, suppose we assume that the only participator channeling sequence in the dynamics is the constant sequence with value χ: A_i always channels with $A_{\chi(i)}$. Then Proposition 1.3 says, in effect, that our stochastic process is a Markov chain whose transition probability from time t_0 to t_0+1 is the kernel $N_{t_0,\chi}(\cdot,\cdot)$. Recall that in **6**–4 we treat a simple case of this artificial situation for $k = 2$. We there assume that the system consists of two participators which channel to each other at each instant of reference time. Thus, of the five elements of $\mathcal{I}(2)$, we assume, artificially, that the only relevant one is χ, where $\chi(1) = 2$.[1]

Although we consider in this book only systems where the number k of participators is fixed, we believe it is unreasonable to build a general theory on the further assumption that the channeling arrangement χ is also fixed. But then on what does χ depend? One might suppose, for example, that each participator comes equipped with a set of channeling affinities, one for each participator in the ensemble. But this, too, seems artificial. For participators are individuated instantaneously by their perspectives, and it is natural to suppose that the channeling affinities depend, at least in part, on these perspectives. This idea suggests that channeling affinity is attached somehow to E, wherein it describes "mutual perceptual accessibility" or "informational conductivity" between pairs of perspectives. In symmetric frameworks the affinity might depend only on the ***difference*** of perspectives, i.e., on the group J.

What form should information on channeling affinities take in order that we may use it to compute the transition probabilities for the markovian dy-

---

[1] We do not use the kernel $N_{t_0,\chi}$ form for the transition probability of the dynamics in **6**–4 although we could have done so. Instead, we there represent the dynamics in a form which is "relativized with respect to the first participator." This means that we are looking at a chain whose states are the displacements (in the group J) of the second participator with respect to the first.

namics in our participator ensemble? If we want to base our computations on formulae like those in Proposition 1.3, we need to know the probabilities, denoted $\tau(e_1, \ldots, e_k; \chi)$, of the various possible χ at time t_0, conditioned by the perspectives of the A_i at t_0, i.e., conditioned by $(e_1, \ldots e_k)$. If we know these $\tau(e_1, \ldots, e_k; \chi)$ we can take the weighted sum

$$\begin{aligned} &N_{t_0}(e_1, \ldots, e_k; \Delta_1 \times \ldots \times \Delta_k) \\ &\quad = \sum_{\chi \in \mathcal{I}(k)} \tau(e_1, \ldots, e_k; \chi) N_{t_0,\chi}(e_1, \ldots, e_k; \Delta_1 \times \ldots \times \Delta_k). \end{aligned} \tag{2.2}$$

$N_{t_0}(e_1, \ldots, e_k; \Delta_1 \times \ldots \times \Delta_k)$ is the probability of a perspective change from $(e_1, \ldots, e_k)$ to $\Delta_1 \times \ldots \times \Delta_k$ from time t_0 to $t_0 + 1$, regardless of the channeling pattern: it is the desired transition probability for the chain in E^k generated by the participator ensemble $A_1, \ldots, A_k$.

One natural way to define the probabilities τ might be to utilize a ***metric*** on E. Intuitively one could define the "elementary" probability that two observers will channel to each other in terms of the distance in this metric between the points of E representing their perspectives. $\tau(e_1, \ldots, e_k; \chi)$ could then be computed in some canonical way using these elementary probabilities. However, since the study of the Markov chains is our primary interest, we simply assume the existence of a τ satisfying certain formal properties. Thus we do not consider the interesting question of the possible relation of τ to other intrinsic data such as metrics on E.

We assume that the τ-distribution is attached to the reflexive framework itself, and not to any particular set of participators. Therefore, τ should be defined for any k, and its expression for various values of k should be consistent. The following definition is a minimal one with these properties:

Definition 2.3. A ***τ-distribution*** is a family $\tau = \{\tau_k\}_{k=1}^{\infty}$ where each τ_k is a markovian kernel on $E^k \times 2^{\mathcal{I}(k)}$,[2] i.e., a map

$$\tau_k \colon E^k \times 2^{\mathcal{I}(k)} \to [0, 1]$$

satisfying the following conditions:

1. $\tau_k(\cdot\,; \chi) \in \mathcal{E}^k$ for all $\chi \subset \mathcal{I}(k)$, and $\tau_k(y_1, \ldots, y_k; \cdot)$ is a probability distribution on $\mathcal{I}(k)$ for all $(y_1, \ldots, y_k) \in E^k$.

[2] This notation means that we are viewing $\mathcal{I}(k)$ as a measurable space with σ-algebra $2^{\mathcal{I}(k)}$.

2. *Consistency condition.* Given $k' < k$, let $S' = \{1, \ldots, k'\}$, $S = \{1, \ldots, k\}$. Then, (with the notation of 1.2 above) for any

$$(y_1, \ldots, y_{k'}, z_{k'+1}, \ldots, z_k) \in E^k,\ \chi \in \mathcal{I}(k'),$$

we have

$$\tau_{k'}(y_1, \ldots, y_{k'};\, \chi) = \frac{\sum_{\chi'' \in \mathcal{I}(S-S')} \tau_k(y_1, \ldots, y_{k'},\, z_{k'+1}, \ldots, z_k;\, \chi \cup \chi'')}{\sum_{\substack{\chi' \in \mathcal{I}(S') \\ \chi'' \in \mathcal{I}(S-S')}} \tau_k(y_1, \ldots, y_{k'},\, z_{k'+1}, \ldots, z_k;\, \chi' \cup \chi'')}.$$

3. *Symmetry conditions.* If our reflexive framework is a symmetric framework, we consider two symmetry conditions on τ corresponding to two notions of equivalence on E^k. First, we define two k-tuples, $y = (y_1, \ldots, y_k)$ and $x = (x_1, \ldots, x_k) \in E^k$, to be *configuration equivalent* if for every i and l satisfying $1 \leq i, l \leq k$ we have $x_i x_l^{-1} = y_i y_l^{-1}$ (notation as in 1.1). We define them to be *translation equivalent* if there exists a $j \in J$ with $x_i = j y_i$, for all $1 \leq i \leq k$. Then
   (i) τ is *configuration symmetric* if $\tau(x;\cdot) = \tau(y;\cdot)$ whenever x and y are configuration equivalent;
   (ii) τ is *translation symmetric* if $\tau(x;\cdot) = \tau(y;\cdot)$ whenever x and y are translation equivalent.

Intuitively, condition 2 states that the probability of any particular channeling among a system S' of observers is not affected by the addition of extra observers to the system as long as there is no "cross-channeling," i.e., as long as one conditions the probability of channeling by those channelings which pair no member of S' with one of the added observers. It is this condition which unites the separate τ_k's for various k.

In condition 3, if J were commutative then configuration equivalence and translation equivalence would be identical. In this case configuration symmetry and translation symmetry would be identical conditions on τ. When J is noncommutative, however, the two conditions are different. In fact the two notions of equivalence on E^k may not even be comparable (in the sense that an equivalence class from one relation is not necessarily a union of equivalence classes from the other).

Except when we wish to emphasize a particular k, we omit the subscript k of the τ; we simply write $\tau(y_1, \ldots, y_k;\, \chi)$.

We do not require $\tau(y_1, \ldots, y_k; \cdot)$ to be invariant under permutations of $y_1, \ldots, y_k$. Thus, the formalism permits the encoding of channeling affinities between specific participators even though such affinities seem naturally attached to the framework itself and not to specific participator ensembles.

Remark 2.4. For each k, consider the product space $E^k \times \mathcal{I}(k)$; denote it $\hat{E}^k$. Let $\hat{\mathcal{E}}^k$ denote the σ-algebra on $\hat{E}^k$ which is generated by $\mathcal{E}^k = \mathcal{E} \otimes \ldots \otimes \mathcal{E}$ and the algebra of all subsets of $\mathcal{I}(k)$. Let $p = \mathrm{pr}_1$ denote projection on the first factor of $\hat{E}^k$, i.e., $p\colon \hat{E}^k \to E^k$. p is measurable, and each fibre of p is a copy of $\mathcal{I}(k)$. As defined in 2.3, the τ-distribution is a kernel on $E^k \times 2^{\mathcal{I}(k)}$. We may also view it as a kernel on $E^k \times \hat{\mathcal{E}}^k$ as follows. Let $A \in \hat{\mathcal{E}}^k$ be a measurable set in $\hat{E}^k$ and let $(e_1, \ldots, e_k) \in E^k$. Define

$$\bar{\tau}(e_1, \ldots, e_k;\, A) = \tau(e_1, \ldots, e_k;\, \mathrm{pr}_2[A \cap p^{-1}\{(e_1, \ldots, e_k)\}]),$$

where $\mathrm{pr}_2\colon \hat{E}^k \to \mathcal{I}(k)$ is projection onto the second factor of $\hat{E}^k$. $\mathrm{pr}_2[A \cap p^{-1}\{(e_1, \ldots, e_k)\}]$ consists of just those channelings χ such that $(e_1, \ldots, e_k;\, \chi)$ is in A.

It is clear from Definition 2.3 that $\bar{\tau}$ is a markovian kernel on $E^k \times \hat{\mathcal{E}}^k$ such that $\bar{\tau}(e_1, \ldots, e_k;\, \cdot)$ is concentrated on $p^{-1}\{(e_1, \ldots, e_k)\}$. In the sequel, we think of τ as this $\bar{\tau}$, using the same symbol τ for both.

3. Augmented dynamics

In this section we discuss the underlying probabilistic framework for the participator dynamics. In section two the τ distribution was motivated by the need to compute the transition probability for the Markov chain whose states are the positions in E of the k participators in a given ensemble. Thus the state space of this chain is E^k, and it is called the (absolute) *position chain* of the dynamics. (The word “absolute” here is sometimes used, in the context of a symmetric framework, to distinguish this chain from other chains which express the positions relative to some fixed observer or to one of the participators; these latter chains have state space J^k or J^{k-1}, respectively.)

Now a knowledge only of the positions of the participators in perspective space is insufficient for many purposes; we need information not only about positions, but also about channelings. The situation at any moment of reference time is most completely described, in our development, by a vector in $\hat{E}^k$ (defined in 2.4 as $E^k \times \mathcal{I}(k)$). We may refer to the elements of $\hat{E}^k$ as being “augmented” by the inclusion of the channeling involution in the description. The stochastic process with state space $\hat{E}^k$ is then the underlying probabilistic framework for all other processes discussed in this chapter and the next one. We refer to the process in $\hat{E}^k$ as the *augmented position chain.*

Before describing the augmented dynamics, we note that there are yet other chains relevant to participator dynamics. For example, the symmetrized perspective space is E^k/S_k (S_k being the symmetric group of permutations of k objects). In this state space we are unconcerned with the identities of the k participators and note only the set of perspectives assumed by them. The corresponding stochastic process is called the "symmetrized position chain." On a different tack, we may to our dynamical situation associate "stopped chains," i.e., chains descended from the reference time chains via ***stopping times.*** For example, we might stop the reference time chains when a given participator A is channeled to, and note the positions of all participators only at such times. Such a chain is clearly derived from the augmented position chain, to the study of which we now turn.

3.1. Let Θ denote the reflexive observer framework $(X, Y, E, S, \pi_\bullet)$. Suppose we are given a τ-distribution on Θ as in Definition 2.3 above. We will describe the ***augmented dynamics of an ensemble of k participators*** $(\xi_1, Q_1(n)), \ldots, (\xi_k, Q_k(n))$ ***on*** Θ. This is a Markov chain indexed by reference time t, with state space $\hat{E}^k = E^k \times \mathcal{I}(k)$; a state of this chain encodes the location in E of each of the participators at a given reference time, as well as specifying the channeling relation among them at that time.

To describe a Markov chain it suffices to give a starting measure on the state space, and a one-step transition probability for each time t. In our present situation this transition probability will be a markovian kernel

$$\hat{N}_t \colon \hat{E}^k \times \hat{\mathcal{E}}^k \to [0, 1].$$

Here $\hat{\mathcal{E}}^k$ denotes the σ-algebra (defined in Remark 2.4) on $\hat{E}^k = E^k \times \mathcal{I}(k)$ generated by all sets of the form $\Delta \times \{\chi\}$ where $\Delta \in \mathcal{E}^k$ and $\chi \in \mathcal{I}(k)$. Thus $\hat{N}_t$ is completely determined once we express

$$\hat{N}_t(e_1, \ldots, e_k, \chi_0; \Delta \times \{\chi_1\})$$

in terms of our given participators. This notation means the following: $\hat{N}_t$ is the probability that at time $t+1$ our k-tuple of participators will have perspectives represented by a point in $\Delta \subset E^k$ and will channel to each other as dictated by an involution χ_1 in $\mathcal{I}(k)$, given that at time t they had perspectives $(e_1, \ldots, e_k)$ and channeled to each other according to $\chi_0 \in \mathcal{I}(k)$.

These considerations suggest the following definition:

Definition 3.2. Let $(e, \chi_0) \in E^k \times \mathcal{I}(k)$, with $e = (e_1, \ldots, e_k)$. Let $\Delta \times \{\chi_1\} \in \hat{\mathcal{E}}^k$. Define

$$\hat{N}_t(e, \chi_0;\, \Delta \times \{\chi_1\}) = \int_\Delta N_{t,\chi_0}(e_1, \ldots, e_k;\, dy_1 \ldots dy_k)\tau(y_1, \ldots, y_k;\, \chi_1),$$

where N_{t,χ_0} is as in Proposition 1.3. In other words

$$\begin{aligned}&\hat{N}_t(e, \chi_0;\, \Delta \times \{\chi_1\})\\ &= \int_\Delta \Big[\prod_{i \in D(\chi_0)} Q_{ie_i}(t) \left(e_{\chi_0(i)}; dy_i\right) \prod_{j \notin D(\chi_0)} \epsilon_{e_j}(dy_j)\Big] \tau(y_1, \ldots, y_k; \chi_1).\end{aligned}$$

If our participators are kinematical, i.e., time independent, then $\hat{N}_t$ is independent of t, and we call it simply $\hat{N}$. In this case the augmented dynamical chain is a *homogeneous* Markov chain with transition probability $\hat{N}$.

Notation 3.3. To stress the dependence of $\hat{N}_t$ on the $Q_i(t)$ and τ, and for other reasons to be discussed later, we sometimes use the notation $\widehat{\langle Q_1(t), \ldots, Q_k(t)\rangle}_\tau$ instead of $\hat{N}_t$. Similarly, if the participators are kinematical we may use the notation $\widehat{\langle Q_1, \ldots, Q_k\rangle}_\tau$ instead of $\hat{N}$.

The action kernels of our k participators together with τ give rise to the transition probabilities of the augmented dynamical chain. Similarly, the initial measures ξ_i of these participators together with τ determine the starting measure of this chain on $E^k \times \mathcal{I}(k)$ as we now describe.

Notation 3.4. Let ξ be a measure on E^k. We denote by ξ_τ the measure on $E^k \times \mathcal{I}(k)$ given by

$$\xi_\tau(\Delta_1 \times \ldots \times \Delta_k \times \{\chi\}) = \int_{\Delta_1 \times \ldots \times \Delta_k} \xi(dy_1 \ldots dy_k)\tau(y_1, \ldots, y_k;\, \chi).$$

If ξ is a probability measure, so is ξ_τ.

Proposition 3.5. Let $p\colon \hat{E}^k = E^k \times \mathcal{I}(k) \to E^k$ be projection on the first factor, i.e., $p = \mathrm{pr}_1$. Then $p_*(\xi_\tau) = \xi$, and $m_p^{\xi_\tau} = \tau$ (using the notation of **2**–1).

Proof. The proof is an exercise in the definition of the τ-distribution (see Remark 2.4) and of a regular conditional probability distribution. The situation is especially simple since the fibres of p are copies of the discrete space $\mathcal{I}(k)$. ∎

Definition 3.6. The starting measure of the augmented dynamical chain of the ensemble of participators $(\xi_1, Q_1(t)), \ldots, (\xi_k, Q_k(t))$ is $(\xi_1 \otimes \ldots \otimes \xi_k)_\tau$.

The interpretation of this measure is straightforward: we assume that the initial positions of the participators are distributed independently, so that

$$(\xi_1 \otimes \ldots \otimes \xi_k)_\tau(\Delta_1 \times \ldots \times \Delta_k \times \{\chi\})$$

is the probability that, at starting time $t = 0$, the k-tuple of perspectives of our participators lie in $\Delta_1 \times \ldots \times \Delta_k \subset E^k$ and channel according to χ in $\mathcal{I}(k)$.

We summarize. On the reflexive framework $\Theta = (X, Y, E, S, \pi_\bullet)$ with given τ-distribution, suppose we have an ensemble of participators $(\xi_1, Q_1(t))$, $\ldots$, $(\xi_k, Q_k(t))$. Associated to this situation is a canonical Markov chain with state space $\hat{E}^k = E^k \times \mathcal{I}(k)$, called the augmented dynamical chain of the participator ensemble.

The state of this chain at time t is given by random variables $y_1(t)$, $\ldots$, $y_k(t)$ (with values in E) and $\chi(t)$ (with values in $\mathcal{I}(k)$). $y_i(t)$ is the perspective of the ith participator, and $\chi(t)$ is the channeling relation among the k participators, at time t. For fixed t these variables are not independent: the dependence of $\chi(t)$ on the $y_i(t)$ is expressed by the τ-distribution. Moreover, since the dynamics is markovian, the dependence of the distribution of the $y_i(t+1)$ and $\chi(t+1)$ on previous values can be expressed entirely in terms of the $y_i(t)$ and $\chi(t)$. This expression is contained in the one-step transition probability at time t, denoted $\widehat{\langle Q_1(t), \ldots, Q_k(t)\rangle}_\tau$ or $\hat{N}_t$ (the precise definition is given in 3.2). The starting measure of the chain is $(\xi_1 \otimes \ldots \otimes \xi_k)_\tau$ (using Notation 3.4).

Notation 3.7. The "base space" of the augmented position chain is the probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ on which the random variables $y_i(t)$, $\chi(t)$ (for all i, t) are defined. Thus the "sample" space $\hat{\Omega}$ is the domain of these random variables; we take it as the space $(\hat{E}^k)^\infty$ of all trajectories $t \to (y_1(t), \ldots, y_k(t); \chi(t))$. If we want to emphasize the number k of participators generating the chain, we write $(\hat{\Omega}^k, \hat{\mathcal{F}}^k, \hat{P}^k)$ instead of just $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$. We write

$$\hat{y}(t) = (\hat{y}_1(t), \ldots, \hat{y}_k(t), \chi(t)),$$

so that for each t, $\hat{y}(t) \colon \hat{\Omega} \to E^k \times \mathcal{I}(k)$. (We follow the usual probabilistic convention of suppressing explicit mention of the sample points $\hat{\omega} \in \hat{\Omega}$ unless necessary.) By our choice of $\hat{\Omega}$ then, $\hat{y}(t)$ is the "t"th coordinate vector of the trajectory. The σ-algebra $\hat{\mathcal{F}}$ is taken to be that generated in $\hat{\Omega}$ by the sequence of random variables $\{\hat{y}(t)\}$. The probability measure $\hat{P}$ on the sample space is developed from the initial measure $(\xi_1 \otimes \ldots \otimes \xi_k)_\tau$ and the transition probabilities $\hat{N}_t$ in canonical fashion. In this sense the augmented position chain is presented as a "canonical" Markov chain[3], with filtration $\{\hat{\mathcal{F}}_n\}$ where $\hat{\mathcal{F}}_n = \sigma(\hat{y}(0), \ldots, \hat{y}(n))$.

4. Augmented dynamics and standard dynamics

Suppose that we are in a reflexive framework Θ with given τ-distribution and that we have an ensemble of participators $(\xi_1, Q_1(t)), \ldots, (\xi_k, Q_k(t))$ as in the previous section. Suppose further that for all χ in $\mathcal{I}(k)$ we know $N_{t,\chi}$, as in Proposition 1.3. (Recall that $N_{t,\chi}$ is the transition probability for the k-tuple of perspectives from time t to time $t+1$ assuming that the particular channeling relation χ occurred at time t.) Then, as in Equation 2.2, we can define the kernel N_t on E^k.

Definition 4.1. For $(e_1, \ldots, e_k) \in E^k$, $\Delta_1 \times \ldots \times \Delta_k \in \mathcal{E}^k$,

$$\begin{aligned} &N_t(e_1, \ldots, e_k;\, \Delta_1 \times \ldots \times \Delta_k) \\ &\quad = \sum_{\chi \in \mathcal{I}(k)} \tau(e_1, \ldots, e_k;\, \chi) N_{t,\chi}(e_1, \ldots, e_k; \Delta_1 \times \ldots \times \Delta_k) \\ &\quad = \sum_{\chi \in \mathcal{I}(k)} \tau(e_1, \ldots, e_k;\, \chi) \prod_{i \in D(\chi)} Q_i(t)(e_{\chi(i)} e_i^{-1};\, \Delta_i e_i^{-1}) \prod_{j \notin D(\chi)} 1_{\Delta_j}(e_j). \end{aligned}$$

[3] See Remark 5.9 of this chapter.

If all our participators are kinematical with fixed action kernels Q_i, then we can omit mention of t.

Notation 4.2. We sometimes use the notation $\langle Q_1(t), \ldots, Q_k(t)\rangle_\tau$ in place of N_t, or $\langle Q_1, \ldots, Q_k\rangle_\tau$ in place of N in the kinematical case.

With notation and hypotheses as above, the following definition is natural:

Definition 4.3. *The standard dynamical chain generated by an ensemble of participators* is the canonical Markov chain with state space E^k, one step transition probabilities $N_t = \langle Q_1(t), \ldots, Q_k(t)\rangle_\tau$, and starting measure $\xi_1 \otimes \ldots \otimes \xi_k$.

Notation 4.4. We denote the base sample space of this standard dynamical chain by Ω, or Ω^k if we want to emphasize the particular value of k. Thus Ω $(= \Omega^k) = (E^k)^\infty$. The chain, then, consists formally of the sequence of random variables $y(t)\colon \Omega \to E^k$ $(t = 0, 1, \ldots)$, where $y(t) = (y_1(t), \ldots, y_k(t))$.

In this section we study certain aspects of the relationship between the augmented dynamical chain and the standard dynamical chain for a given ensemble of participators. The following diagram summarizes the basic setup:

$$\begin{array}{ccccc} & & \text{augmented chain} & \hat{E}^k & \hat{N}_t \\ \hat{\Omega} = & (E^k \times \mathcal{I}(k))^\infty & \xrightarrow{\hat{y}(t)=(y_1(t),\ldots,y_k(t),\chi(t))} & E^k \times \mathcal{I}(k) & \\ & \Big\downarrow{\scriptstyle p'} & & \Big\downarrow{\scriptstyle p=\mathrm{pr}_1} & \\ \Omega & (E^k)^\infty & \xrightarrow{y(t)=(y_1(t),\ldots,y_k(t))} & E^k & N_t \\ & & \text{standard chain} & & \end{array} \tag{4.5}$$

p' is induced by p. To exercise the notation, let $\hat{\omega}$ be an element of $\hat{\Omega}$. We can view $\hat{\omega}$ as a sequence of elements of $E^k \times \mathcal{I}(k)$ indexed by t, i.e.,

$$\hat{\omega} = \{(e_1(t), \ldots, e_k(t), \chi(t))\}_{t=1}^\infty.$$

Then

$$\begin{aligned}\hat{y}(t)(\hat{\omega}) &= (e_1(t), \ldots, e_k(t), \chi(t)) \\ &= (y_1(t)(\hat{\omega}), \ldots, y_k(t)(\hat{\omega}), \chi(t)(\hat{\omega})). \\ p'(\hat{\omega}) &= \{(e_1(t), \ldots, e_k(t))\}_{t=1}^{\infty}, \\ y(t)p'(\hat{\omega}) &= (e_1(t), \ldots, e_k(t)), \quad \text{etc.}\end{aligned}$$

The top and bottom rows of diagram 4.5 represent the augmented and standard dynamical chains respectively, for which the one step transition probabilities are, respectively, $\hat{N}_t = \widehat{\langle Q_1(t), \ldots, Q_k(t)\rangle}_\tau$ and $N_t = \langle Q_1(t), \ldots, Q_k(t)\rangle_\tau$.

Now there is an abstract characterization of the structural relationship between $\hat{N}_t$ and N_t, which does not follow merely from the simple relationship between the state spaces of the two chains. It can be understood in terms of general operations on kernels which we now introduce.

The first part of the following definition merely recalls the notion of "pushdown" of a measure, introduced in **2**–1. The second part then generalizes this notion to kernels.

Definition 4.6. Let $(U, \mathcal{U})$ and $(V, \mathcal{V})$ be measurable spaces and let $h: U \to V$ be a measurable function.

(i) If μ is a measure on U, the ***pushdown of*** μ ***by*** h is the measure $h_*\mu$ on V, given by

$$h_*\mu(A) = \mu(h^{-1}(A)), \quad A \in \mathcal{V}.$$

Alternatively, for any measurable $g: V \to \mathbf{R}$,

$$\int_V (h_*\mu)(dv)g(v) = \int_U \mu(du)(g \circ h)(u).$$

(ii) If M is a kernel on U, the ***pushdown of*** M ***by*** h is the kernel h_*M on $U \times \mathcal{V}$, given by

$$(h_*M)(u, A) = M(u, h^{-1}(A)), \quad A \in \mathcal{V}.$$

Again, we may restate this in terms of operations on functions:

$$(h_*Mg)(u) = \int_U M(u, du')\; g \circ h(u'), \quad g \in \mathcal{V}.$$

If μ is a probability measure, so is $h_*\mu$; if M is markovian, so is h_*M. The notion of composition of a measure and a kernel, or of two kernels, was introduced in **6**–1. We generalize it here.

Definition 4.7. Let $(U,\mathcal{U})$, $(V,\mathcal{V})$ and $(W,\mathcal{W})$ be measurable spaces. Let K be a kernel on $U \times \mathcal{V}$.

(i) If μ is a measure on U, the measure μK on V is defined by

$$\mu K(A) = \int_U \mu(du)K(u,A), \quad A \in \mathcal{V}.$$

(ii) If L is a kernel on $W \times \mathcal{U}$, the composition LK is the kernel on $W \times \mathcal{V}$ defined by

$$LK(w,A) = \int_U L(w,du)K(u,A), \quad A \in \mathcal{V}.$$

As in 4.7, we may easily write down the effect of these compositions on functions $g: V \to \mathbf{R}$. Also, if μ is a probability measure and K and L are markovian kernels, then μK is a probability measure and KL is markovian.

Combining these definitions we have the following:

Definition 4.8. Let $(U,\mathcal{U})$ and $(V,\mathcal{V})$ be measurable spaces with $h: U \to V$ measurable. Let M be a kernel on U and L a kernel on $V \times \mathcal{U}$. The *L-pushdown of M by h* is then the kernel $h_*^L M$ on V, defined by

$$\begin{aligned}(h_*^L M)(v,A) &= (L(h_*M))(v,A) \\ &= \int_U L(v,du)M(u,h^{-1}(A)), \quad u \in U, A \in \mathcal{V}.\end{aligned}$$

We wish to use this construction to relate the kernel $\hat{N}_t$ on $\hat{E}^k$ (in place of M on U) to the kernel N_t on E^k (where E^k replaces V). The role of h is played by $p = \mathrm{pr}_1$ on $\hat{E}^k$, while that of L is played by $\bar{\tau}$ as in 2.4. As mentioned in 2.4 we will, however, write just τ in place of $\bar{\tau}$, viewing the τ-distribution as a kernel on $E^k \times \hat{\mathcal{E}}^k$. Pictorially,

$$
\begin{array}{rll}
 & & \hat{N}_t \colon \hat{E}^k \times \hat{\mathcal{E}}^k \to [0,\ 1] \\
E^k \times \mathcal{I}(k) = & \hat{E}^k & \tau \colon E^k \times \hat{\mathcal{E}}^k \to [0,\ 1] \\
 & \big\downarrow \mathrm{pr}_1 = p & \\
 & E^k & p_*^\tau \hat{N}_t \colon E^k \times \mathcal{E}^k \to [0,\ 1]
\end{array}
$$

Using Definition 4.8 we get the kernel $p_*^\tau \hat{N}_t$ on E^k:

$$p_*^\tau \hat{N}_t(e, A) = \int_{\hat{E}^k} \tau(e, d\hat{e}) p_* \hat{N}_t(\hat{e}, A).$$

Intuitively, the above pushdown consists in averaging the values $\hat{N}_t(\cdot, p^{-1}(A))$ with respect to the measure $\tau(e, \cdot)$. Now the measure $\tau(e, \cdot)$ is concentrated on the fibre $p^{-1}\{e\}$; recall that this fibre may be viewed as a copy of $\mathcal{I}(k)$. Thus $p_*^\tau \hat{N}_t(e, A)$ is an expectation of the values $p_* \hat{N}_t(e, \chi_0; A)$ with respective weights $\tau(e, \chi_0)$. These values can, in turn, be related to the objects N_{t,χ_0} (Proposition 1.3) as follows. We claim that, for any $A \in \mathcal{E}^k$,

$$p_* \hat{N}_t(e, \chi_0; A) = N_{t,\chi_0}(e; A). \tag{4.9}$$

For (suppressing the subscript t)

$$
\begin{aligned}
p_* \hat{N}(e, \chi_0; A) &= \hat{N}(e, \chi_0; p^{-1}(A)) \\
&= \hat{N}(e, \chi_0; \bigcup_{\chi \in \mathcal{I}(k)} A \times \{\chi\}) \\
&= \sum_{\chi \in \mathcal{I}(k)} \hat{N}(e, \chi_0; A \times \{\chi\}) \\
&= \sum_{\chi \in \mathcal{I}(k)} \int_A N_{\chi_0}(e; de') \tau(e'; \chi) \\
&= \int_A N_{\chi_0}(e; de') \tau(e'; \mathcal{I}(k)) \\
&= \int_A N_{\chi_0}(e; de') = N_{\chi_0}(e;\ A).
\end{aligned}
$$

Taking the expectation of this over all $\chi_0 \in \mathcal{I}(k)$ with respect to the measure $\tau(e; \cdot)$, we recover the transition probability N_t:

Proposition 4.10. $p_*^\tau \hat{N}_t = N_t$.

Proof. $\tau(e,\cdot)$ is concentrated on $p^{-1}\{e\}$, which is a copy of $\mathcal{I}(k)$, so that the τ-pushdown of $\hat{N}_t$ via p is a sum:

$$\begin{aligned} &p_*^\tau(\hat{N})(e;A)\\ &\qquad = \sum_{\chi_0\in\mathcal{I}(k)} \tau(e;\chi_0)N_{\chi_0}(e;\, A) \quad \text{by (4.9)}\\ &\qquad = N(e;\, A) \quad \text{by Definition 4.1.} \end{aligned}$$ ∎

The previous proposition describes the “algebraic” relationship between the kernels $\hat{N}$ and N. However, this by itself does not completely clarify the *probabilistic* relationship between the augmented and the standard chains. To achieve this further understanding, we first recall from chapter two the notion of regular conditional probability distribution, expressed in terms of the algebra of pushdowns and compositions. Using the notation of 4.7, we may state the criteria for a kernel K on $V\times\mathcal{U}$ to be a version of the rcpd of a measure μ on U with respect to h:

(i) For $h_*\mu$-almost all v, $K(v,\cdot)$ is a probability measure concentrated on $h^{-1}\{v\}$.

(ii)
$$\mu = (h_*\mu)\cdot K. \tag{4.11}$$

In this case we write $K = m_h^\mu$ and

$$\mu = (h_*\mu)\cdot m_h^\mu. \tag{4.12}$$

Now consider the *measures* $\hat{N}_t(y,\chi_0;\cdot)$ on $\hat{E}^k$ for fixed y and χ_0. Their rcpd decomposition is, if it exists,

$$\begin{aligned} \hat{N}_t(y,\chi_0;\cdot) &= [p_*\hat{N}_t(y,\chi_0;\cdot)][m_p^{\hat{N}_t(y,\chi_0;\cdot)}]\\ &= [N_{t,\chi_0}(y;\cdot)][m_p^{\hat{N}_t(y,\chi_0;\cdot)}] \end{aligned} \tag{4.13}$$

by 4.9. The measures $N_{t,\chi_0}(y;\cdot)$ in general differ for different values of y and χ_0. However, the “orthogonal” parts of the decomposition do not depend on y and χ_0.

Proposition 4.14. For any $\hat{y}\in\hat{E}^k$, $m_p^{\hat{N}(\hat{y};\,\cdot)} = \tau$. (As in Proposition 4.8, for simplicity of notation we have suppressed the subscript t in $\hat{N}_t$; and we continue to view τ as a kernel on $E^k\times\hat{\mathcal{E}}^k\to[0,\,1]$ as in 2.4.)

Proof. In view of 4.11 and 4.13, we must show that, for $\hat{y} = (y, \chi_0) \in \hat{E}^k$,

$$\begin{aligned}\hat{N}(\hat{y};\cdot) &= \hat{N}(y,\chi_0;\cdot)\\ &= \int_{E^k} N_{\chi_0}(y;dw)\int_{p^{-1}\{w\}} \tau(w,\cdot)\end{aligned}$$

by 4.8. It is enough to verify this formula applied to sets of the form $A \times \{\chi_1\}$, with $A \in \mathcal{E}^k$ and $\chi_1 \in \mathcal{I}(k)$, since any measurable set in $\hat{\mathcal{E}}$ is a finite union of such sets. Thus, we are to show

$$\hat{N}(\hat{y};\, A \times \{\chi_1\}) = \int_{E^k} N_{\chi_0}(y;\, dw) \int_{p^{-1}\{w\}} \tau(w;\, d\hat{z}) 1_{A\times\{\chi_1\}}(\hat{z}).$$

Recall now that $\tau(w;\, d\hat{z}) = \tau(w;\, \chi)d\chi$ where $d\chi$ denotes counting measure on $p^{-1}\{w\} = \{w\} \times \mathcal{I}(k)$. Thus the right hand side of the last equation may be written as

$$\begin{aligned}&\int_{E^k} N_{\chi_0}(y;\, dw) \sum_{\chi\in\mathcal{I}(k)} \tau(w;\, \chi) 1_{A\times\{\chi_1\}}(w,\, \chi)\\ &\qquad = \int_A N_{\chi_0}(y;\, dw)\tau(w;\, \chi_1).\end{aligned}$$

Thus our original equation is seen to be

$$\hat{N}(\hat{y};\, A \times \{\chi_1\}) = \int_A N_{\chi_0}(y;\, dw)\tau(w;\, \chi_1)$$

which is the same as Definition 3.2. ∎

In the next section we consider the general setting in which the probabilistic significance of 4.9 and 4.14 is clarified.

5. Descent of Markov chains

We now consider the concept of *descent* of a Markov chain. Suppose we have a Markov chain whose base space is the probability space $(B, \mathcal{B}, \rho)$, whose filtration is $\{\mathcal{G}_t\}$, and whose state space is $(U, \mathcal{U})$. The random variables of the chain are denoted by $u_t\colon B \to U$, $t = 0, 1, 2, \ldots$. Now let $h\colon (U,\mathcal{U}) \to (V,\mathcal{V})$ be a measurable function, and let $v_t = h \circ u_t$.

$$\begin{array}{ccc} B & \xrightarrow{u_t} & U \\ & v_t \searrow & \Big\downarrow h \\ & & V \end{array}$$

The sequence $\{v_t\}$, along with $(B, \mathcal{B}, \rho)$ and the natural filtrations $\{\sigma(v_0, \ldots, v_t)\}$, forms a stochastic process.

Terminology 5.1. The Markov chain $\{u_t\}$ *descends via* h if the stochastic process $\{v_t\}$ is also a Markov chain.

The distribution of v_t is induced by h from the distribution of u_t: If $A \in \mathcal{V}$ then $\rho(v_t \in A) = \rho(u_t \in h^{-1}(A))$. In particular, if the starting measure of the chain $\{u_t\}$ is ν, that of $\{v_t\}$ is $h_*\nu$. A well-known condition for the descent of a chain is expressed in the following definition and theorem:

Definition 5.2. Given a bimeasurable $h\colon U \to V$ and a kernel M on U, we will say that M is h-*respectful* if, for any $A \in \mathcal{V}$, $M(u_1, h^{-1}(A)) = M(u_2, h^{-1}(A))$ whenever $h(u_1) = h(u_2)$. Associated to such an M is a kernel on V, denoted $R_h M$ and defined by

$$R_h M(v, A) = M(u, h^{-1}(A)) = h_* M(u, A)$$

for any $u \in h^{-1}\{v\}$.

Remark 5.3. The bimeasurability of h ensures that $R_h M$ is indeed a kernel on V. The h-respectfulness of M is equivalent to the condition that $h_* M(\cdot, A)$, defined in 4.6, is constant on fibres of h: we have

$$h_* M(u, A) = R_h M(h(u), A).$$

Example 5.4. Suppose $f\colon U \to [0,\infty]$ is measurable. This gives us a kernel I_f on U defined as follows:

$$I_f(u, du') = f(u)\epsilon_u(du').$$

These are the simplest kernels; in particular, f could be 1_C for a measurable subset C of U (in which case we write I_C for I_{1_C}). The kernel I_f, then, is h-respectful if and only if f is measurable with respect to the σ-algebra $h^*\mathcal{V}$ of h, i.e., if and only if there is some measurable function g on V such that $f = g \circ h$.[4] For then, if $A \in \mathcal{V}$,

$$\begin{aligned} I_f(u, h^{-1}(A)) &= f(u)1_{h^{-1}(A)}(u) \\ &= g(h(u))1_A(h(u)) \end{aligned}$$

so that respectfulness holds. Furthermore,

$$R_h I_f \equiv R_h I_{g\circ h} = I_g,$$

where I_g is thought of as a kernel on V. In the special case where $f = 1_C$, the condition for respectfulness amounts to saying that $C = h^{-1}(C')$ for some subset C' of V; the measurability of C' being a consequence of the bimeasurability of h. In this instance

$$R_h I_C \equiv R_h I_{h^{-1}(C')} = I_{C'}.$$

Respectfulness allows us to prune the state space from U to V, a space which more efficiently carries the essential information of the kernel.

Theorem 5.5. With notation as above, suppose that $\{u_t\}$ is a Markov chain with respect to the family $\{\mathcal{G}_t\}$ of subσ-algebras of $\mathcal{B}$ on B. Suppose that the one step transition probabilities M_t of the chain are h-respectful. Then the chain $\{u_t\}$ descends via h; the one-step transition probabilities $R_h M_t$ of the chain $\{v_t\}$ are given, for $v \in V$ and $A \in \mathcal{V}$, by

$$R_h M_t(v, A) = M_t(u, h^{-1}(A)),$$

[4] See Parthasarathy (1977, Proposition 44.1) for a proof of this statement.

where u is ***any*** element in $h^{-1}\{v\}$. Moreover, $\{v_t\}$ is a Markov chain with respect to the same sequence $\{\mathcal{G}_t\}$ of σ-algebras on B (and not just the sequence $\{\sigma(v_0,\ldots,v_t)\}$).

The condition of the h-respectfulness of the $\{M_t\}$ is sufficient for the chain $\{u_t\}$ to descend via h, but it is far from necessary. In fact, we now state a different sufficient condition. In this case the chain descends in a slightly weaker sense: The $\{v_t\}$ is now a Markov chain only with respect to the subσ-algebras $\{\sigma(v_0,\ldots,v_t)\}$ of $\mathcal{B}$, and only when the measure on B is of a special type. It is worth mentioning that the two conditions on the $\{M_t\}$ appear to be completely independent, having in common only that they are both sufficient for the descent of the chain.

As before, let $(U,\mathcal{U})$ and $(V,\mathcal{V})$ be measurable spaces and let $h\colon U \to V$ be a measurable function. Suppose we are given a family $\{M_t\}_{t=0,1,2,\ldots}$ of kernels on U. In particular, for each t and each $u \in U$, $M_t(u,\cdot)$ is a measure for $\mathcal{U}$. In principle we may then consider the rcpd's of these various measures with respect to h, i.e., we may consider the kernels

$$m_h^{M_t(u,\cdot)}\colon V \times \mathcal{U} \to [0,1].$$

These rcpd's may not exist in the most general situation, but they will exist, for example, if $(U,\mathcal{U})$ and $(V,\mathcal{V})$ are standard Borel spaces.

Definition 5.6. The family of kernels $M = \{M_t\}$ is ***h-decomposable*** if there exists a single kernel m on $V \times \mathcal{U}$ which is, for each $u \in U$ and $t \geq 0$, a version of the rcpd of $M_t(u,\cdot)$ with respect to h. We will speak of m as a "common rcpd" of M.

We also speak of the h-decomposability of a ***single*** kernel, with the obvious meaning.

In case of h-decomposability, the kernels $h_*^m M_t$ on V defined in Definition 4.8 are naturally associated to M_t; we will denote them also as $D_h M_t$ when there is no confusion regarding the version of common rcpd being used. Then

$$D_h M_t(v,A) = h_*^m M_t(v,A) = \int_U m(v,du) M_t(u,h^{-1}(A)).$$

Example 5.7. The family of kernels $\{\hat{N}_t\}$, which are the one-step transition probabilities of an augmented dynamical chain, is p-decomposable, where as usual

$$p\colon \hat{E}^k \to E^k$$

is projection. Indeed, by Proposition 4.14 all of the rcpd's of the measures $\hat{N}_t(\hat{y}, \cdot)$ are equal to τ.

One might ask under what conditions the kernels I_f, defined in 5.4, are h-decomposable. A calculation shows that this happens only in a somewhat trivial case. Namely, the support of f must lie within the set of those $u \in U$ through which the fibre of h is the singleton $\{u\}$ itself. In this case a common rcpd m of I_f may be described as follows. Suppose $\bar{h}\colon V \to U$ is a "measurable section of the fibre bundle defined by $h\colon U \to V$." That is, $\bar{h}(v) \in h^{-1}\{v\}$ for all $v \in V$. Then a version of the common rcpd of the measures $I_f(u, \cdot)$ is given by

$$m(v, du') = \epsilon_{\bar{h}(v)}(du').$$

The function f, supported as it is only within the singleton fibres of h, is an h-measurable function. As such, there is some measurable function g on V such that $f = g \circ h$. A computation then shows that in fact

$$D_h I_f = R_h I_f = I_g.$$

However, we will see in Proposition **8**–4.15 that the h-decomposability of a kernel K implies the h-decomposability of the product $I_f K$ for any $f \geq 0$.

In general, suppose we have a family $\{M_t\}$ of markovian kernels which we are interpreting as the one-step transition probabilities of a Markov chain $\{u_t\}$ with state space U. Thus $u_t\colon B \to U$, for $t = 0, 1, 2, \ldots$, is a random variable defined on the base probability sample space B. In this case $M_t(u_t, \cdot)$ is the conditional distribution of u_{t+1} given u_t, and so the h-decomposability of $\{M_t\}$ has the following interpretation: for all t, the conditional expectation of u_{t+1} given $h(u_{t+1})$ is independent of u_t. In statistical terminology, we can say that for each t the statistic h of the random variable u_{t+1} is ***sufficient*** for the "parameter" u_t.

From this point of view the p-decomposability of the $\{\hat{N}_t\}$, and, indeed, the conclusion of Proposition 4.14 itself, becomes intuitively clear given the definition of the τ-distribution. Namely, the fibres of $p\colon \hat{E}^k \to E^k$ are all copies of $\mathcal{I}(k)$. At any time t the measure on $\mathcal{I}(k)$ which describes the conditional distribution of the augmented state $\hat{e}_{t+1} = (e_{t+1}, \chi_{t+1})$, given $p(\hat{e}_t) = e_t$, is $\tau(p(\hat{e}_t), \cdot) = \tau(e_t, \cdot)$. And this depends only on the value e_t in E^k, and not on χ_t per se.

Theorem 5.8. [5] Let $(B, \mathcal{B}, \rho)$ be a probability space and let $(U, \mathcal{U})$ and $(V, \mathcal{V})$ be measure spaces. Let $\{u_t\}$, $t = 0, 1, 2, \ldots$ be a Markov chain in U with base B, and with one-step transition probabilities given by the family of kernels $\{M_t\}$. Let $h: U \to V$ be measurable and let $v_t = h \circ u_t$. Suppose that the family $\{M_t\}$ is h-decomposable; let ψ denote their common rcpd with respect to h. Let ν denote the starting measure of the chain, i.e., $\nu = u_{0*}(\rho)$, the distribution of u_0. Suppose that ψ is also the rcpd of ν with respect to h. Then $\{v_t\}$ is a Markov chain in V with base $(B, \mathcal{B}, \rho)$, transition probabilities $h_*^{\psi}(M_t)$, and initial measure $h_*\nu$.

Remark 5.9. The terminology means that $\{u_t\}$ is a Markov chain with respect to the increasing family $\{\sigma(u_0, \ldots, u_t)\}$ of subσ-algebras of $\mathcal{B}$, while $\{v_t\}$ is a Markov chain with respect to $\{\sigma(v_0, \ldots, v_t)\}$.

Before turning to the proof of Theorem 5.8 we will first recall some basic facts about the *canonical chain.*

Let $\Omega = U \times U \times \ldots$; a typical element of Ω will be denoted $\omega = (x_0, x_1, \ldots)$. Let $\mathcal{F}$ denote the σ-algebra of Ω generated by "measurable rectangles," i.e., by sets of the form $A_0 \times A_1 \times \ldots$, where the A_i are in $\mathcal{U}$ and only finitely many of them are different from U. Given the kernels $\{M_t\}$ and the starting measure ν on U, we construct a measure M_ν on $(\Omega, \mathcal{F})$ as follows:

$$\begin{aligned} &M_\nu(A_0 \times A_1 \times \ldots) \\ &\quad = \int_{A_0} \nu(dx_0) \int_{A_1} M_0(x_0, dx_1) \ldots \int_{A_n} M_{n-1}(x_{n-1}, dx_n) \ldots . \end{aligned} \tag{5.10}$$

Let $X_t: \Omega \to U$ denote projection onto the tth factor, $t = 0, 1, \ldots$. Let $\mathcal{F}_t = \sigma(X_0, \ldots, X_t)$ be the smallest σ-algebra on Ω with respect to which the $X_0, \ldots, X_t$ are measurable. $\mathcal{F}_t$ is then the subσ-algebra of $\mathcal{F}$ generated by those measurable rectangles of the form $A_1 \times \ldots \times A_t \times U \times U \times \ldots$. Then we have

Proposition 5.11. With these hypotheses and notation:

1. The distribution of X_0 is ν.
2. $\{X_t\}$ is a Markov chain with base $(\Omega, \mathcal{F}, M_\nu)$ (with respect to the subσ-algebras $\mathcal{F}_t$) with one-step transition probabilities $\{M_t\}$. It is called the *canonical chain* for those M_t with the given starting measure ν.

[5] We are indebted to D. Revuz for informing us that a related result for continuous time may be found in Pitman and Rogers (1981).

3. Suppose $(B, \mathcal{B}, \rho)$ is a probability space and $\{u_t\}: B \to U$ is a Markov chain with the same transition probabilities $\{M_t\}$ and starting measure $u_{0*}(\rho) = \nu$. Then there is a unique (ρ-a.s.) measurable function $\phi: B \to \Omega$ such that $X_t \circ \phi = u_t$ (ρ-a.s.) and $\phi_*(\rho) = M_\nu$. This is called *the universal property of the canonical chain.*

Proof of Theorem 5.8. In view of the universal property of the canonical chain, we may assume that $(B, \mathcal{B}, \rho) = (\Omega, \mathcal{F}, M_\nu)$. We will still use the notation $\{u_t\}$ to denote the random variables defining the chain; $u_t: \Omega \to U$ is now projection on the tth factor. We still denote $v_t = h \circ u_t$. Let $\mathcal{G}_t$ now denote the subσ-algebra $\sigma(v_0, \ldots, v_t)$ of $\mathcal{F}$. Concretely, $\mathcal{G}_t$ is generated by all measurable rectangles of the form $h^{-1}(A_0) \times \ldots \times h^{-1}(A_t) \times U \times U \times \ldots$, $A_i \in \mathcal{V}$. We will temporarily denote $h^\psi_* M_t$ by K_t. Pictorially,

$$\begin{array}{ccccc} M_\nu & \Omega & \xrightarrow{u_t} & U & \qquad \nu,\ M_t \\ & v_t & \searrow & \Big\downarrow h & \\ & & & V & \quad K_t = h^\psi_* M_t = D_h M_t \end{array}$$

Under the assumption that all the rcpd's of the $M_t(u, \cdot)$, as well as that of ν, are equal to ψ, we are going to show that $\{v_t\}$ is a Markov chain with one step transition probabilities $\{K_t\}$. Thus we must show that for any $t \geq 1$ and any $\mathcal{V}$-measurable function f on V,

$$(K_{t-1}f)(v_{t-1}) = E[f(v_t)|\mathcal{G}_{t-1}] \quad M_\nu - a.s. \tag{5.12}$$

where $E = E_{M_\nu}$ denotes expectation with respect to the measure M_ν on Ω. To prove this, since $(K_{t-1}f)(v_{t-1})$ is clearly $\mathcal{G}_{t-1}$-measurable, it is enough to show that for any $A \in \mathcal{G}_{t-1}$ of the form $A = h^{-1}(A_0) \times \ldots \times h^{-1}(A_{t-1}) \times U \times U \times \ldots$,

$$\int_A M_\nu(d\omega)\,(K_{t-1}f)\,(v_{t-1}(\omega)) = \int_A M_\nu(d\omega) f\,(v_t(\omega))\,. \tag{5.13}$$

Now $f(v_t(\omega)) = (f \circ h)(u_t(\omega))$. Since $A \in \mathcal{G}_{t-1} \subset \mathcal{F}_{t-1}$, the right side of 5.13 may be written

$$\int_A M_\nu(d\omega)(f \circ h)(u_t(\omega)) = \int_A M_\nu(d\omega) E[(f \circ h)(u_t)|\mathcal{F}_{t-1}](\omega).$$

But since the Markov chain $\{u_t\}$ has transition probabilities $\{M_t\}$, 5.13 becomes

$$\int_A M_\nu(d\omega)\left(K_{t-1}f\right)\left(h\circ u_{t-1}(\omega)\right)=\int_A M_\nu(d\omega)M_{t-1}\left(f\circ h\right)\left(u_{t-1}(\omega)\right). \tag{5.14}$$

In view of the definition of the measure M_ν (as in (5.10)) and the set A, the integrals in (5.14) can be written as iterated integrals in the variables $u_0,\ldots,u_{t-1}$ successively. Since the integrands involve only u_{t-1}, to show the integrals are equal it suffices to consider only the last iteration on each side, i.e., it suffices to show

$$\int_{h^{-1}(A_{t-1})} M_{t-2}(u_{t-2},du_{t-1})\left(K_{t-1}f\right)\left(h(u_{t-1})\right)$$
$$=\int_{h^{-1}(A_{t-1})} M_{t-2}(u_{t-2},du_{t-1})(M_{t-1}f\circ h)(u_{t-1})$$

where $A_{t-1}\in\mathcal{V}$ and $u_{t-1},\ u_{t-2}\in U$ are arbitrary. Note that if $t-1=0$ we must be careful to interpret the symbol $M_{t-2}(u_{t-2},\cdot)$ (which is then, a priori, meaningless) to be the measure $\nu(\cdot)$. In other words, we must prove

$$\int_{h^{-1}(C)} M_{t-2}(u,dx)\left(K_{t-1}f\right)\left(h(x)\right)=\int_{h^{-1}(C)} M_{t-2}(u,dx)M_{t-1}\left(f\circ h\right)(x), \tag{5.15}$$

where $C\in\mathcal{V}$, $u\in U$, x is a variable on U, and $M_{t-2}(u,\cdot)$ is defined to be $\nu(\cdot)$ if $t-1=0$.

We now evaluate the left side of 5.15. Recalling Definition 4.6(ii), we see that

$$\begin{aligned}\int_{h^{-1}(C)} & M_{t-2}(u,dx)\left(K_{t-1}f\right)\left(h(x)\right)\\ &=\int M_{t-2}(u,dx)1_C(h(x))(K_{t-1}f)(h(x))\\ &=\int h_*M_{t-2}(u,dv)\left(K_{t-1}f\right)(v)1_C(v)\\ &=\int h_*M_{t-2}(u,dv)\left(h_*^{\psi}M_{t-1}f\right)(v)1_C(v)\end{aligned}$$

by definition of K_{t-1}. By Definition 4.8, this is the same as

$$\int_C h_*M_{t-2}(u,dv)\int_{1_C(v)}\psi(v,dx)M_{t-1}(f\circ h)(x).$$

Now we use the fact that, for any u, ψ is the rcpd of $M_{t-2}(u,\cdot)$ with respect to h. Since $\psi(v,\cdot)$ is supported on $h^{-1}\{v\}$, $1_C(v)\psi(v,dx)$ is the same as

$$\psi(v,dx)1_{h^{-1}(C)}(x).$$

Thus the above integral is

$$\int_{h^{-1}(C)} M_{t-2}(u,dx)M_{t-1}(f\circ h)(x).$$

But this is just the right hand side of (5.15) above. ∎

The conditions of h-respectfulness and h-decomposability thus allow us to compute the transition probabilities of the descended chain. In case of descent via h, the distribution of $v_t = h(u_t)$ is given by $h_*(\nu M_0 M_1 \dots M_t)$; if the descent is respectful or decomposable we may explicitly express this as $h_*\nu \cdot R_h M_0 \cdot R_h M_1 \dots R_h M_t$ or $h_*\nu \cdot D_h M_0 \cdot D_h M_1 \dots D_h M_t$ respectively. Further discussion of the descent conditions may be found in **8**–3.

6. Summary of formulae

A. *Pushdowns*

$$h_*\mu(A) = \mu(h^{-1}(A)) \tag{4.6(i)}$$

$$h_*M(u,A) = M(u,h^{-1}(A)) \tag{4.6(ii)}$$

$$h_*^L M(v,A) = (L\cdot h_*M)(v,A) \tag{4.8}$$

B. *Descents*

(1) Respectful

$$R_h M(v,A) = h_*M(u,A), \quad u\in h^{-1}\{v\} \tag{5.2}$$

(2) Decomposable

$$D_h M(v,A) = m\cdot h_*M, \quad m = m_h^{M(u,\cdot)},\ \forall u \tag{5.6}$$

C. *Initial measures*

(1) Standard

$$\begin{aligned}\xi(A) &= (\xi_1 \otimes \ldots \otimes \xi_k)(A) \\ &= \int \xi_1(dy_1)\ldots\xi_k(dy_k)1_A(y_1,\ldots,y_k), \quad A \in \mathcal{E}^k\end{aligned} \tag{4.3}$$

(2) Augmented

$$\xi_\tau(A \times \{\chi\}) = \int_{E^k} \xi(dy)1_A(y)\tau(y;\chi), \quad A \in \mathcal{E}^k,\ \chi \in \mathcal{I}(k) \tag{3.4}$$

D. *Transition probabilities*

(1) Fixed channeling

$$N_{t,\chi}(e,A) = \int_A \prod_{i\in D(\chi)} Q_{i,e_i}(t)(e_{\chi(i)};dy_i) \prod_{j\notin D(\chi)} \epsilon_{e_j}(dy_j) \tag{1.3}$$

(2) Augmented

$$\begin{aligned}\hat{N}_t(e,\chi_0;A\times\{\chi_1\}) = \int_{E^k} N_{t,\chi_0}(e;dy)1_A(y)\tau(y;\chi_1),& \\ e \in E^k;\ A \in \mathcal{E}^k;\ \chi_0,\chi_1 \in \mathcal{I}(k)&\end{aligned} \tag{3.2}$$

(3) Standard

$$N_t(e;A) = \sum_{\chi\in\mathcal{I}(k)} \tau(e;\chi)N_{t,\chi}(e;A), \quad e \in E^k,\ A \in \mathcal{E}^k \tag{4.1}$$

$$N_t = p_*^\tau \hat{N}_t = D_p\hat{N}_t \tag{4.14}$$