Random Measurable Selections

Jean Goubault-Larrecq[1]   Roberto Segala[2]

[1] ENS Cachan   goubault@lsv.ens-cachan.fr
[2] Università di Verona   roberto.segala@univr.it

Abstract. We make the first steps towards showing a general "random-ness for free" theorem for stochastic automata. The goal of such theorems is to replace randomized schedulers by averages of pure schedulers. Here, we explore the case of measurable multifunctions and their measurable selections. This involves constructing probability measures on the measurable space of measurable selections of a given measurable multifunction, which seems to be a fairly novel problem. We then extend this to the case of IT automata, namely, non-deterministic (infinite) automata with a history-dependent transition relation. Throughout, we strive to make our assumptions minimal.

1   Introduction

This paper grew out of an attempt at proving a "randomness for free" type theorem [5] for stochastic automata [4]. We present the first steps in this direction.

A stochastic automaton is a transition system on a measurable space Q of *states*. When in state $q \in Q$, we have access to a set $\theta(q)$ of fireable transitions, from which we choose non-deterministically. A transition is a pair (a, μ) of an *action* a, from a fixed measurable space L, and of a probability measure μ on Q. Once we have chosen such a transition, we pick the next state q' at random with respect to μ, and proceed. There are two ways to resolve the non-deterministic choice of a transition $(a, \mu) \in \theta(q)$. A *pure scheduler* σ is a function that maps each path $w = q_0 a_1 q_1 \cdots a_n q_n$ of states and actions seen so far to an element $\sigma(w)$ of $\theta(q_n)$ (or to a special termination constant $\perp$). A *randomized scheduler* η instead maps w to a (sub)probability measure concentrated on $\theta(q_n)$, thereby drawing the transition at random as well. In each case, given a scheduler η, the stochastic automaton behaves as a purely probabilistic transition system, and one can define the probability $P_\eta(\mathcal{E})$ that the automaton will follow a path that lies in the measurable set $\mathcal{E}$.

The "randomness for free" question we envision is as follows: given a randomized scheduler η, and a measurable set $\mathcal{E}$ of paths, can we find two *pure schedulers* σ^-, σ^+ such that $P_{\sigma^-}(\mathcal{E}) \leq P_\eta(\mathcal{E}) \leq P_{\sigma^+}(\mathcal{E})$? This has a number of important applications, and was solved positively by Chatterjee, Doyen *et al.* [5, Section 4], in the case that Q and L are finite. In general, we consider the following more general question: given a randomized scheduler η, and a measurable *payoff* function h from paths to $\mathbb{R}^+$, can we find two pure schedulers

σ^-, σ^+ such that $\int_\omega h(\omega)dP_{\sigma^-} \leq \int_\omega h(\omega)dP_\eta \leq \int_\omega h(\omega)dP_{\sigma^+}$? This includes the previous question, by taking the characteristic function $\chi_\mathcal{E}$ of $\mathcal{E}$ for h.

Beware that one cannot reasonably ask for the existence of a pure scheduler σ such that $P_\sigma(\mathcal{E}) = P_\eta(\mathcal{E})$. For example, let $Q = \{0, 1, 2\}$, $L = \{*\}$, $\theta(0) = \{t_1 = (*, \delta_1), t_2 = (*, \delta_2)\}$ (where δ_x is the Dirac mass at x), $\theta(1) = \theta(2) = \emptyset$. There are only two pure schedulers, one that reaches state 1 with probability 1, while the other reaches 2 with probability 1. But one can reach 1 with *arbitrary probability* p using the randomized scheduler $\eta(0) = p\delta_{t_1} + (1-p)\delta_{t_2}$.

While Chatterjee, Doyen *et al.* are concerned with finite state and action spaces, infinite spaces are useful as well: in modeling timed probabilistic and non-deterministic transition systems [7, Example 1.1], or devices that interact with the physical world [6, 1], where each state includes information about time, position, impulse and other real-valued data, for example. Considering his publication record (see for example [16] on labeled Markov processes, or [10] on Markov decision processes), Prakash would probably be the first to approve us.

Our initial aim was to prove such "randomness for free" theorems for general stochastic automata. During the year 2008, we found very complex arguments that proved only very weak versions of what we were after. We were on the wrong path. On the opposite, Chatterjee, Doyen *et al.* [5, Section 4] used a simple idea: draw pure schedulers σ themselves at random, with respect to some measure ϖ, designed so that $P_\eta(\mathcal{E}) = \int_\sigma P_\sigma(\mathcal{E})d\varpi$. The claim then follows by standard integration arguments (Fact 1 below).

The probability measure ϖ on pure schedulers has an intuitive description, too: we merely choose the transition $\sigma(w)$ at random with respect to probability $\eta(w)$, where η is our given randomized scheduler, for each finite path w, independently. Now this is the source of typical measure-theoretic complications. First, we must force our pure schedulers to be measurable. Otherwise, $\int_\sigma P_\sigma(\mathcal{E})d\varpi$, and in fact already P_σ, makes no sense. Second, we need to make clear what the intended σ-algebra is on the space of all pure schedulers. If we don't have any, nothing of the above makes any sense either. Third, what does it mean to draw $\sigma(w)$ for each w *independently*? The sheer fact that σ is measurable must enforce at least some moderate amount of non-independence.

Chatterjee, Doyen *et al.* did not need to address these issues: on finite state and action spaces, *all* pure schedulers are measurable, and the problems above disappear. Going to infinite spaces of states and actions requires extra effort.

While we were writing this paper, we soon realized that we would have to choose between: (1) solving the full question, by applying all the required measure-theoretic clout if necessary, possibly making it so complex that nobody would understand it; or (2) solving a few restricted cases, showing a few elegant proof tricks along the way. It should be clear that (2) was a better choice. We hope that Prakash will appreciate the techniques, if not the results.

That is, we shall be content to solve the problem in the special case of *IT automata*, namely, stochastic automata with a trivial action space, no option for termination, and no random choice at all: randomness will be induced by the randomized scheduler only. We shall deal with the general case in another paper.

The plan of the paper is as follows. We recapitulate some required basic measure theory in Section 2. In Section 3, we define a σ-algebra on the set $Sel(F)$ of *measurable selections* of a given multifunction F—these are just what we have called pure schedulers, for the transition relation F of a stochastic automaton without probabilistic choice; and we show that any randomized scheduler g defines a canonical probability measure ϖ_g on $Sel(F)$ such that, for every x, drawing a point at random with probability $g(x)$ gives the same result as drawing a measurable selection f with probability ϖ_g and computing $f(x)$ (Proposition 1). This is the most important construction of the paper. In Section 4, we extend this result from random measurable selections to random measurable pure schedulers of IT automata. Although the setting looks extremely similar, there is no hope of reusing the previous result. Instead, we use similar proof techniques, but with a more complex implementation. We conclude in Section 5.

2 Basics on Measure Theory

A *σ-algebra* on a set X is a family of subsets that is closed under complement and countable unions. We shall write $\overline{A}$ for the complement of A in X. A pair $X = (|X|, \Sigma_X)$ of a set $|X|$ and a σ-algebra Σ_X on $|X|$ is a *measurable space*, and the elements of Σ_X are called the *measurable subsets* of X. We shall sometimes write X instead of $|X|$ to avoid pedantry.

Given any family $\mathcal{F}$ of subsets of a set A, there is a smallest σ-algebra that contains the elements of $\mathcal{F}$. This is called the σ-algebra *generated by* $\mathcal{F}$. We shall sometimes refer to the elements of $\mathcal{F}$ as the *basic* measurable subsets of this σ-algebra, despite some ambiguity. One example is $\mathbb{R}$ with its *Borel σ-algebra*, generated by intervals. Another one is the *product* $X_1 \times X_2$ of two measurable spaces is $(|X_1| \times |X_2|, \Sigma_{X_1} \otimes \Sigma_{X_2})$, whose basic measurable subsets are the *rectangles* $E_1 \times E_2$, $E_1 \in \Sigma_{X_1}$, $E_2 \in \Sigma_{X_2}$. In general, the σ-algebra on the product of an arbitrary family of measurable spaces $(X_i)_{i \in I}$ is the one generated by the subsets $\pi_i^{-1}(E)$ where $E \in \Sigma_{X_i}$, $i \in I$, and $\pi_i \colon \prod_{i \in I} |X_i| \to |X_i|$ is the usual projection onto coordinate i.

The *coproduct* $X_1 + X_2$ of two measurable spaces X_1, X_2 is simpler: $|X_1 + X_2|$ is the disjoint union of $|X_1|$ and $|X_2|$, and $\Sigma_{X_1+X_2}$ consists of unions $E_1 \cup E_2$ of a measurable subset E_1 of X_1 and of a measurable subset E_2 of X_2. This construction generalizes to countable coproducts $\sum_{n \in \mathbb{N}} X_n$ in the obvious way.

A *measurable* map $f \colon X \to Y$ is one such that $f^{-1}(E) \in \Sigma_X$ for every $E \in \Sigma_Y$. If $\mathcal{F}$ generates Σ_Y, it is enough to check that $f^{-1}(E) \in \Sigma_X$ for every E in $\mathcal{F}$ to establish the measurability of f.

A *measure* μ on X is a map from Σ_X to $\mathbb{R}^+ \cup \{+\infty\}$ that is countably additive ($\mu(\emptyset) = 0$, and $\mu(\bigcup_{n \in \mathbb{N}} E_n) = \sum_{n \in \mathbb{N}} \mu(E_n)$ for every countable family of disjoint measurable subsets E_n). A *probability* measure is one such that $\mu(X) = 1$. The *Dirac mass* at x, δ_x, is the probability measure defined by $\delta_x(E) = 1$ if $x \in E$, 0 otherwise.

A measure μ is *concentrated on* a measurable subset A of X if and only if $\mu(X \smallsetminus A) = 0$. For example, if X is finite and $\Sigma_X = \mathbb{P}(|X|)$, then $\mu = \sum_{x \in X} a_x \delta_x$

is concentrated on $\{x \in X \mid a_x \neq 0\}$. Any subset A (even non-measurable) of $|X|$ gives rise to a measurable *subspace*, again written A, with $\Sigma_A = \{A \cap B \mid B \in \Sigma_X\}$. If A is measurable, one can define the *restriction* $\mu_{|A}$ of μ to the subspace A, by $\mu_{|A}(B) = \mu(B)$ for every $B \in \Sigma_A$. If μ is a probability measure that is concentrated on A, then $\mu_{|A}$ is also a probability measure.

There is a standard notion of integral of measurable maps $h \colon X \to \mathbb{R}^+$ with respect to a measure μ on X, which we write $\int_{x \in X} h(x)d\mu$. Other notations found in the literature are $\int_{x \in X} h(x)\mu(dx)$ or $\langle h, \mu \rangle$. We shall also use the notation $\int_{x \in X} d\mu\, h(x)$, especially when $h(x)$ is a long formula, as in $\int_{x_1 \in E_1} d\mu_1 \int_{x_2 \in E_2} d\mu_2\, h(x_1, x_2)$. Writing χ_E for the characteristic map of a measurable subset E of X, $\int_{x \in X} \chi_E(x)d\mu$ equals the measure $\mu(E)$ of E.

Given a map $f \colon A \to |X|$ (not necessarily measurable) where X is a measurable space and A is a set, the family $(f^{-1}(E))_{E \in \Sigma_X}$ is a σ-algebra on A, called the σ-algebra *induced by f on A*. When f is the inclusion map of a subset A of $|X|$, we retrieve the subspace σ-algebra Σ_A.

If μ is concentrated on a measurable subset A of X, then $\int_{x \in X} g(x)d\mu = \int_{x \in A} g(x)d\mu_{|A}$, where A is considered as a subspace of X on the right-hand side.

We write $\mathcal{P}(X)$ for the space of all probability measures on X, with the *weak σ-algebra*, generated by the subsets $[E > r] = \{\mu \in |\mathcal{P}(X)| \mid \mu(E) > r\}$. One can equate $\mu \in \mathcal{P}(X)$ with a vector of real numbers $(\mu(E))_{E \in \Sigma_X}$, i.e., with an element of the measurable space $\mathbb{R}^{\Sigma_X}$. The weak σ-algebra is nothing else than the σ-algebra induced by the inclusion of $|\mathcal{P}(X)|$ into the product $\mathbb{R}^{\Sigma_X}$.

Given a measurable map $f \colon X \to Y$, and a measure μ on X, the formula $f[\mu](E') = \mu(f^{-1}(E'))$ defines a measure $f[\mu]$ on Y, called the *image measure* of μ by f. For any measurable $h \colon Y \to \mathbb{R}^+$, the following *change of variables* formula holds [2, Theorem 16.13]:

$$\int_{y \in Y} h(y)df[\mu] = \int_{x \in X} (h \circ f)(x)d\mu. \tag{1}$$

More trivially, the function $\mathfrak{e}_X \colon X \to \mathcal{P}(X)$ that sends x to the Dirac mass δ_x is measurable. These facts assemble to define the so-called Giry monad [11, Section 1], of which $\mathfrak{e}$ is the unit. (More precisely, one of the two Giry monads.) Prakash stressed the importance of this monad in [14]—probably the one paper that popularized it.

Finally, we shall use the following well-known fact near the end of the paper.

Fact 1 *For every integrable map $h \colon X \to \mathbb{R}$ on a measurable space X, for every $a \in \mathbb{R}$, if $\int_{x \in X} h(x)d\mu = a$ for some probability measure μ on X, then there are points $x^-, x^+ \in |X|$ such that $h(x^-) \leq a \leq h(x^+)$.*

Indeed, if x^- did not exist, say, then $h(x) > a$ for every $x \in |X|$. Let $A_n = h^{-1}(a + 1/n, +\infty)$ for every non-zero natural number n: so $|X| = \bigcup_{n \geq 1} A_n$. Since $h(x) \geq a$ for every $x \in |X|$ and $h(x) \geq a + 1/n$ if additionally $x \in A_n$, $a = \int_{x \in X} h(x)d\mu \geq a + 1/n\, \mu(A_n)$, so $\mu(A_n) = 0$. A consequence of σ-additivity is that the measure of the union of a countable chain of measurable subsets is the sup of their measures, so $1 = \mu(|X|) = \sup_{n=1}^{+\infty} \mu(A_n) = 0$: contradiction.

Carathéodory's measure extension theorem. The following measure extension theorem, due to Carathéodory, was singled out as "a very useful type of theorem" by Prakash [15, Theorem 18]; see also [2, Theorem 11.3].

A *semiring* $\mathcal{A}$ on a set Ω is a collection of subsets of Ω that contains the empty set, is closed under binary intersections, and such that the difference $A \smallsetminus B$ of any two sets $A, B \in \Omega$ can be written as a finite union of elements of $\mathcal{A}$. A map μ is *countably subadditive* on $\mathcal{A}$ if and only if for every countable disjoint family of elements A_n of $\mathcal{A}$ whose union A is in $\mathcal{A}$, $\mu(A) \leq \sum_{n \in \mathbb{N}} \mu(A_n)$.

Theorem 1 (Carathéodory). *Let Ω be a set, and $\mathcal{A}$ be a semiring on Ω. Every function $\mu \colon \mathcal{A} \to [0, +\infty]$ such that $\mu(\emptyset) = 0$, that is finitely additive and countably subadditive on $\mathcal{A}$, extends to a measure on the σ-algebra generated by $\mathcal{A}$. In particular, this is so if $\mu(\emptyset) = 0$ and μ is countably additive on $\mathcal{A}$.*

A typical application is $\Omega = \mathbb{R}$, $\mathcal{A}$ is the semiring of all half-closed intervals $[a, b)$, and $\mu[a, b) = b - a$, leading to Lebesgue measure.

The Ionescu-Tulcea Theorem. Let Q_n, $n \in \mathbb{N}$, be countably many measurable spaces, and assume countably many measurable maps $g_n \colon \prod_{i=0}^{n-1} Q_i \to \mathcal{P}(Q_n)$. One can think of Q_n as the space of all possible states of a probabilistic transition system at (discrete) time $n \in \mathbb{N}$. Given that at time n we have gone through states $q_0 \in Q_0$, $q_1 \in Q_1$, ..., $q_{n-1} \in Q_{n-1}$, $g_n(q_0, q_1, \cdots, q_{n-1})$ is a probability distribution along which we draw the next state q_n. The following Ionescu-Tulcea Theorem states that these data define a unique probability measure on infinite paths $q_0, q_1, \cdots, q_{n-1}, \cdots$:

Theorem 2 (Ionescu-Tulcea). *Let Q_n, $n \in \mathbb{N}$, be measurable spaces, and $g_* = (g_n \colon \prod_{i=0}^{n-1} Q_i \to \mathcal{P}(Q_n))_{n \geq 1}$ be measurable maps. For every $q_0 \in Q_0$, there is a unique probability measure $P_{g_*}(q_0)$ on $\prod_{i=0}^{+\infty} Q_i$ such that:*

$$
P_{g_*}(q_0)(\prod_{i=0}^{n} E_i \times \prod_{i=n+1}^{+\infty} |Q_i|) = \chi_{E_0}(q_0) \int_{q_1 \in E_1} dg_1(q_0) \int_{q_2 \in E_2} dg_2(q_0 q_1) \cdots \int_{q_n \in E_n} dg_n(q_0 q_1 \cdots q_{n-1}).
$$
(2)

Moreover, P_{g_*} defines a measurable map from Q_0 to $\mathcal{P}(\prod_{i=0}^{+\infty} Q_i)$.

We consider tuples $(q_0, q_1, q_2, \cdots, q_{n-1})$ as words, and accordingly write them as $q_0 q_1 q_2 \cdots q_{n-1}$. The notation $\int_{q_i \in E_i} d\mu_i \, h(q_i)$ (where $\mu_i = g_i(q_0 q_1 \cdots q_{i-1})$ above) stands for $\int_{q_i \in Q_i} d\mu_i \, \chi_{E_i}(q_i) h(q_i)$, and the rightmost integral in Theorem 2 is an integral of the constant 1, which is standardly omitted—i.e., the rightmost integral is $\int_{q_n \in Q_n} \chi_{E_n}(q_n) dg_n(q_0 q_1 q_2 \cdots q_{n-1}) = g_n(q_0 q_1 q_2 \cdots q_{n-1})(E_n)$.

There are several small variations on the Ionescu-Tulcea Theorem. Our version is Giry's [11, Theorem 3], except for the fact that Giry considers more general ordinal-indexed sequences of measurable spaces. We will not require that.

The following is needed in the proof of Lemma 2. Lemma 2 looks perfectly obvious, yet requires some effort to prove. Measure theory is full of these.

Lemma 1. *Under the assumptions of Theorem 2, for every $n \in \mathbb{N}$, and for every measurable map $h\colon \prod_{i=0}^{n} Q_i \to \mathbb{R}^+$, for every $q \in Q_0$,*

$$\int_{q_0 q_1 \cdots \in \prod_{i=0}^{+\infty} Q_i} h(q_0 q_1 \cdots q_n) dP_{g_*}(q) \tag{3}$$
$$= \int_{q_1 \in Q_1} dg_1(q) \int_{q_2 \in Q_2} dg_2(qq_1) \cdots \int_{q_n \in Q_n} dg_n(qq_1 q_2 \cdots q_{n-1}) h(qq_1 q_2 \cdots q_{n-1} q_n).$$

Proof. This is true for functions h of the form $\chi_{\prod_{i=0}^{n} E_i}$, $E_i \in \Sigma_{Q_i}$, as one can check by using (2). Let S be the set of measurable subsets $\mathcal{E}$ of $\prod_{i=0}^{n} Q_i$ such that (3) holds for $h = \chi_{\mathcal{E}}$, i.e., such that $P_{g_*}(q)(\mathcal{E}) = \int_{q_1 \in Q_1} dg_1(q) \int_{q_2 \in Q_2} dg_2(qq_1) \cdots$ $\int_{q_n \in Q_n} dg_n(qq_1 q_2 \cdots q_{n-1}) \chi_{\mathcal{E}}(qq_1 q_2 \cdots q_{n-1} q_n)$. S contains all the rectangles, is closed under complements (using $\chi_{\overline{\mathcal{E}}} = 1 - \chi_{\mathcal{E}}$), and under countable disjoint unions. For the latter, consider countably disjoint elements $\mathcal{E}_m$, $m \in \mathbb{N}$, of S, let $\mathcal{E} = \bigcup_{m \in \mathbb{N}} \mathcal{E}_m$, and realize that $\chi_{\mathcal{E}} = \sup_{m \in \mathbb{N}} \sum_{k=0}^{m} \chi_{\mathcal{E}_k}$. The Monotone Convergence Theorem [2, Theorem 16.2] states that integrals of non-negative real functions commute with pointwise suprema of countable chains, so $P_{g_*}(q)(\mathcal{E}) = \sup_{m \in \mathbb{N}} \sum_{k=0}^{m} \int_{q_1 \in Q_1} dg_1(q) \int_{q_2 \in Q_2} dg_2(qq_1) \cdots \int_{q_n \in Q_n} dg_n(qq_1 q_2 \cdots q_{n-1}) \chi_{\mathcal{E}_k}(qq_1 q_2 \cdots q_{n-1} q_n) = \sum_{m=0}^{+\infty} P_{g_*}(q)(\mathcal{E}_m)$. It follows that S is a σ-algebra containing the rectangles, and therefore contains $\Sigma_{\prod_{i=0}^{n} Q_i}$.

It follows easily that (3) holds for step functions h, i.e., when h is of the form $\sum_{k=0}^{m} a_k \chi_{\mathcal{E}_k}$, $m \in \mathbb{N}$, $a_k \in \mathbb{R}^+$, $\mathcal{E}_k$ measurable. Since every measurable map $h\colon \prod_{i=0}^{n} Q_i \to \mathbb{R}^+$ is the pointwise supremum of a countable chain of step functions (namely $h_m = \sum_{k=1}^{m 2^m} k/2^m \chi_{h^{-1}(k/2^m, +\infty)}$, $m \in \mathbb{N}$), (3) follows by the Monotone Convergence Theorem. $\qquad \square$

Consider now any family of measurable subsets $\mathcal{E}_n$ of $\prod_{i=0}^{n-1} Q_i \times Q_n$, $n \geq 1$, and assume that for all q_0, q_1, $\ldots$, q_{n-1}, $g_n(q_0 q_1 \cdots q_{n-1})$ draws q_n at random so that $q_0 q_1 \cdots q_{n-1} q_n$ is in $\mathcal{E}_n$. It seems obvious that what we shall get in the end is an infinite path $q_0 q_1 \cdots q_n \cdots$ such that every finite prefix $q_0 q_1 \cdots q_n$ is in $\mathcal{E}_n$. This actually needs a bit of proof. Given a measurable subset E of a product $A \times B$, and $a \in |A|$, the *vertical cut* $E_{|a}$ is the set $\{b \in |B| \mid (a, b) \in E\}$. This is measurable as soon as E is [2, Theorem 18.1 (i)].

Lemma 2. *Under the assumptions of Theorem 2, let $\mathcal{E}_n$ be measurable subsets of $\prod_{i=0}^{n-1} Q_i \times Q_n$, $n \geq 1$, and assume that for all $q_0 q_1 \cdots q_{n-1} \in \prod_{i=0}^{n-1} Q_i$, $g_n(q_0 q_1 \cdots q_{n-1})$ is concentrated on $(\mathcal{E}_n)_{|q_0 q_1 \cdots q_{n-1}}$.*

For every $q \in Q_0$, $P_{g_*}(q)$ is concentrated on the set $Path_{\mathcal{E}}$ of infinite paths whose finites prefixes $q_0 q_1 \cdots q_n$ are in $\mathcal{E}_n$ for every $n \geq 1$. If additionally $\{q\}$ is measurable in Q_0, then $P_{g_*}(q)$ is concentrated on the set $Path_{\mathcal{E}}(q)$ of those infinite paths in $Path_{\mathcal{E}}$ such that $q_0 = q$.

Proof. First, $Path_{\mathcal{E}}$ is measurable, as a countable intersection of measurable subsets $\mathcal{E}_n \times \prod_{i=n+1}^{+\infty} |Q_i|$. Since $g_n(q_0 q_1 \cdots q_{n-1})$ is concentrated on $(\mathcal{E}_n)_{|q_0 q_1 \cdots q_{n-1}}$, and the complement of the latter in $|Q_n|$ is $(\overline{\mathcal{E}_n})_{|q_0 q_1 \cdots q_{n-1}}$, $\int_{q_n \in Q_n} \chi_{(\overline{\mathcal{E}_n})_{|q_0 q_1 \cdots q_{n-1}}}$

$dg_n(q_0q_1\cdots q_{n-1}) = g_n(q_0q_1\cdots q_{n-1})((\overline{\mathcal{E}_n})_{|q_0q_1\cdots q_{n-1}}) = 0$. By taking $h = \chi_{\overline{\mathcal{E}_n}}$ in (3), and realizing that $\chi_{\overline{\mathcal{E}_n}}(q_0q_1\cdots q_{n-1}q_n) = \chi_{(\overline{\mathcal{E}_n})_{|q_0q_1\cdots q_{n-1}}}(q_n)$, we obtain that $\int_{q_0q_1\cdots\in\prod_{i=0}^{+\infty} Q_i} \chi_{\overline{\mathcal{E}_n}}(q_0q_1\cdots q_n)dP_{g_*}(q) = 0$. In other words, the $P_{g_*}(q)$-measure of the complement of $\mathcal{E}_n \times \prod_{i=n+1}^{+\infty}|Q_i|$ is zero. As a consequence of σ-additivity, the union of these complements when n ranges over $\mathbb{N}$ has measure that is bounded by the sum of their measures, namely 0. So the complement of $Path_{\mathcal{E}}$ has $P_{g_*}(q)$-measure 0.

For the second claim, if $\{q\}$ is measurable, then $P_{g_*}(q)((|Q_0|\smallsetminus\{q\})\times\prod_{i=1}^{+\infty}|Q_i|)$ is equal to 0 by (2). The measure of the complement of $Path_{\mathcal{E}}(q) = Path_{\mathcal{E}} \cap (\{q\} \times \prod_{i=1}^{+\infty}|Q_i|)$ therefore also has $P_{g_*}(q)$-measure 0. $\qquad\square$

3 Drawing Measurable Selections at Random

Before we go to the more complicated case of schedulers, we illustrate our basic technique on random choice of *measurable selections* of a multifunction. We believe this has independent interest.

A *multifunction* from a set A to a set B is a map F from A to $\mathbb{P}^*(B)$, the non-empty powerset of B. We say that F is *measurable* if and only if its graph $\mathrm{Gr}\, F = \{(x,y) \mid y \in F(x)\}$ is a measurable subset of $X \times Y$. This is one of the many possible notions of measurability for relations, see [12]. The set $F(x)$ is exactly the vertical cut $(\mathrm{Gr}\, F)_{|x}$, showing that for a measurable multifunction, $F(x)$ is a (non-empty) measurable subset of B (see comment before Lemma 2).

A *selection* for a multifunction F is a map $f\colon A \to B$ such that $f(x) \in F(x)$ for every $x \in A$. Every multifunction has a selection: this is the Axiom of Choice. In measure theory, we would like f to be measurable as well. Theorems guaranteeing the existence of measurable selections for certain multifunctions are called *measurable selection theorems*. There are many of them (see Wagner [18], or Doberkat [8]), but one should remember that measurable multifunctions do not have measurable selections in general: Blackwell showed that there is a multifunction from $[0,1]$ to Baire space $\mathbb{N}^{\mathbb{N}}$ whose graph is closed (hence measurable) but has no measurable selection [3] (see also Example 5.1.7 of [17]).

Given two measurable spaces X and Y, let us write $\langle X \to Y\rangle$ for the space of all measurable maps from X to Y, with the *weak σ-algebra*. The latter is by definition the subspace σ-algebra, induced by the inclusion of $|\langle X \to Y\rangle|$ into the product space $Y^{|X|}$. In other words, this is the smallest σ-algebra that makes the maps $\varphi \in \langle X \to Y\rangle \mapsto \varphi(x)$ measurable, for every $x \in |X|$.

More generally, given a multifunction $F\colon |X| \to \mathbb{P}^*(|Y|)$, we also consider the subspace $Sel(F)$ of $\langle X \to Y\rangle$ of all measurable selections of F, with the induced σ-algebra. (Beware that $Sel(F)$ need not be a measurable subset of $\langle X \to Y\rangle$.) We again call the latter the *weak σ-algebra*, on this subset. In each case, the weak σ-algebra is generated by subsets that we write $[x \to E]$, with $x \in |X|$ and $E \in \Sigma_Y$, and defined as those measurable functions, resp. those measurable selections of F, that map x into E.

Assume now a measurable map $g\colon X \to \mathcal{P}(Y)$ such that, for every $x \in |X|$, $g(x)$ is concentrated on $F(x)$. For each $x \in |X|$, pick an element $f(x)$ in $F(x)$ with

probability $g(x)$. The function f is a selection of F, but will not be measurable in general. Can we pick f at random so that f *is measurable* and $f(x)$ is drawn with probability $g(x)$? This is the question we answer in the affirmative here.

The problem looks similar to the construction of Wiener measure, a model of Brownian motion, where we would like to draw a map from $\mathbb{R}$ to some topological space at random, and this map should be continuous [2, Section 37]; or to the construction of Skorokhod's J1 topology, which allows one to make sense of random càdlàg functions. Our solution will be simpler, though: measurability is easier to enforce than continuity or being càdlàg.

One can explain the problem in terms of *independence* [2, Section 5]. Let us remind the reader that independence is not *pairwise* independence. Consider for example two independent random bits b_1 and b_2, and the random variable $b_3 = b_1 \oplus b_2$, where $\oplus$ is exclusive-or. These random variables are pairwise independent, meaning that any pair among them is formed of independent random variables. However, they are *not* independent, since given the value of any two, one obtains the third one in a deterministic way. In our case, if we are to draw a *measurable* map f at random, then the random infinite tuple $(f(x))_{x \in |X|}$ cannot be a collection of independent random variables. However, the results below essentially say that we can choose f *measurable* at random, in such a way that all *countable* sequences $(f(x_n))_{n \in \mathbb{N}}$ are independent.

A general way to draw several values at random, independently, is to draw them with respect to a product measure. The following says that product measures exists not only for finite products but also for countable products of *probability* measures. This is well-known, and can even be extended to uncountable products: this is the Łomnick-Ulam Theorem [13, Corollary 5.18].

Lemma 3. *Let μ_n be probability measures on the measurable spaces X_n, $n \in \mathbb{N}$. There is a unique probability measure μ, written $\bigotimes_{n \in \mathbb{N}} \mu_n$, on $\prod_{n \in \mathbb{N}} X_n$ such that $\mu(\bigcap_{i \in I} \pi_i^{-1}(E_i)) = \prod_{i \in I} \mu_i(E_i)$ for every finite subset I of $\mathbb{N}$, and all measurable subsets E_i of X_i, $i \in I$.*

Proof. Apply Ionescu-Tulcea's Theorem 2 to $Q_0 = \{*\}$, $Q_{n+1} = X_n$, let g_n be the constant map $g_n(q_0 q_1 \cdots q_{n-1}) = \mu_n$, and note that for every finite set I, $\bigcap_{i \in I} \pi_i^{-1}(E_i)$ is just the product $\prod_{i=0}^{n} E_i \times \prod_{i=n+1}^{+\infty} |Q_i|$, for some n large enough, and where we extend the notation E_i to denote $|Q_i|$ for $i \notin I$. $\square$

We shall use the following general technique to construct measurable maps.

Lemma 4 (Patching). *Let X, Y be measurable spaces, $(E_i)_{i \in I}$ be a countable partition of $|X|$ in measurable subsets ($I \subseteq \mathbb{N}$), and $(f_i)_{i \in I}$ be a matching family of measurable maps from E_i to Y. The* patch *$f \colon X \to Y$, defined as mapping every $x \in E_i$ to $f_i(x)$, is a measurable map.*

Proof. Categorically, this follows from the fact that X is the coproduct $\sum_{i \in I} E_i$. Alternatively, $f^{-1}(E) = \bigcup_{i \in I}(f_i^{-1}(E) \cap E_i)$ is measurable as soon as E is. $\square$

As an application, we show that a measurable multifunction F that has a measurable selection must have *plenty* of measurable selections. Precisely, we can fix their values, arbitrarily, at countably many arguments:

Lemma 5. *Let X, Y be measurable spaces, $F\colon X \to \mathbb{P}^*(Y)$ be a measurable multifunction with a measurable selection f, $(x_i)_{i\in I}$ be countably many points in X ($I \subseteq \mathbb{N}$), and y_i be an element of $F(x_i)$ for every $i \in I$. Write $\vec{\mathbf{y}}$ for $(y_i)_{i\in I}$.*

There is a measurable selection $f_{\vec{\mathbf{y}}}$ of F such that $f_{\vec{\mathbf{y}}}(x_i) = y_i$ for every $i \in I$. Moreover, we can choose $f_{\vec{\mathbf{y}}}$ in such a way that the mapping $\vec{\mathbf{y}} \mapsto f_{\vec{\mathbf{y}}}$ is itself a measurable map from $\prod_{i\in I} F(x_i)$ to $Sel(F)$.

Proof. Without loss of generality, assume that I is $\mathbb{N}$, or an initial segment $\{0, 1, \ldots, n-1\}$ of $\mathbb{N}$. Similarly to vertical cuts, we may define horizontal cuts of $\mathrm{Gr}(F)$ at y, namely $\{x \in X \mid y \in F(x)\}$, and they are measurable as well. Write $F^{-1}(y)$ for such a vertical cut. Define E_i, for each $i \in I$, as $F^{-1}(y_i) \smallsetminus \bigcup_{j=0}^{i-1} F^{-1}(y_j)$. Together with $E_\infty = |X| \smallsetminus \bigcup_{i\in I} E_i$, they form a partition of $|X|$ in measurable subsets. Define f_i as the constant map on E_i equal to y_i for $i \in I$, and f_∞ as the restriction of f to E_∞, then form their patch $f_{\vec{\mathbf{y}}}$, using Lemma 4. It is plain to see that $f_{\vec{\mathbf{y}}}$ is a selection of F, and $f_{\vec{\mathbf{y}}}$ is measurable.

To show that $\vec{\mathbf{y}} \mapsto f_{\vec{\mathbf{y}}}$ is itself measurable, we must show that the set A of tuples $\vec{\mathbf{y}}$ such that $f_{\vec{\mathbf{y}}} \in [x \to E]$ is measurable, for $x \in |X|$ and $E \in \Sigma_Y$. For convenience, write $E_i(\vec{\mathbf{y}})$ for the set we called E_i above, and similarly with E_∞. Let E_i' be the set of tuples $\vec{\mathbf{y}} \in \prod_{i\in I} F(x_i)$ such that $y_i \in F(x)$ and $y_j \notin F(x)$ for every j, $0 \le j < i$. E_i' is measurable since $F(x)$ is measurable: E_i' is just the rectangle $\prod_{j=0}^{i-1}(F(x_j) \smallsetminus F(x)) \times (F(x_i) \cap F(x)) \times \prod_{j\in I, j>i} F(x_j)$. Also, $\vec{\mathbf{y}} \in E_i'$ if and only if $x \in E_i(\vec{\mathbf{y}})$. Write π_i for ith projection. Since $f_{\vec{\mathbf{y}}}(x) \in E$ if and only if there is an $i \in I$ such that $x \in E_i(\vec{\mathbf{y}})$ and $y_i \in E$, or for every $i \in I$, $x \notin E_i(\vec{\mathbf{y}})$ and $f(x) \in E$, it follows that $A = (\bigcup_{i\in I} E_i' \cap \pi_i^{-1}(E))$ if $f(x) \notin E$, and $A = (\bigcup_{i\in I} E_i' \cap \pi_i^{-1}(E)) \cup (\bigcap_{i\in I} \overline{E_i'})$ otherwise. In any case, A is measurable. $\square$

Theorem 3 is the keystone of our construction, and allows us to provide foundations to the notion of a random measurable selection.

Theorem 3. *Let X, Y be two measurable spaces, and $F\colon X \to \mathbb{P}^*(Y)$ be a measurable multifunction with a measurable selection. Let also $g\colon X \to \mathcal{P}(Y)$ be a measurable map such that, for every $x \in X$, $g(x)$ is concentrated on $F(x)$.*

There is a unique probability measure ϖ_g on the space $Sel(F)$ of measurable selections of F such that $\varpi_g(\bigcap_{i=1}^n [x_i \to E_i]) = \prod_{i=1}^n g(x_i)(E_i)$ for every finite collection of pairwise distinct points $(x_i)_{1\le i\le n}$ of X and of measurable subsets $(E_i)_{1\le i\le n}$ of Y.

Before we prove it, we note the following consequence. Proposition 1 can also be seen as a partial implementation of the Chatterjee-Doyen-Gimbert-Henzinger idea of the introduction in the Markovian case: given a (Markovian) randomized scheduler g, draw (Markovian) pure schedulers f at random so that $f(x)$ is drawn with probability $g(x)$.

Proposition 1. *Under the assumptions of Theorem 3, let $h\colon Y \to \mathbb{R}^+$ be a measurable map, and x be a point of X, then:*

$$\int_{f\in Sel(F)} h(f(x)) d\varpi_g = \int_{y\in Y} h(y) dg(x).$$

Proof. Let α_x be the measurable map $f \in Sel(F) \mapsto f(x)$. By the change of variables formula (1), $\int_{f \in Sel(F)} h(f(x))d\varpi_g = \int_{y \in Y} h(y)d\alpha_x[\varpi_g]$. Now note that $\alpha_x[\varpi_g](E) = \varpi_g(\alpha_x^{-1}(E)) = \varpi_g([x \to E]) = g(x)(E)$, so $\alpha_x[\varpi_g] = g(x)$. $\qquad\square$

Proof (of Theorem 3). We use Carathéodory's measure extension Theorem 1. Let $\mathcal{A}$ be the semiring of subsets of the form $\bigcap_{i=1}^{n}[x_i \to E_i]$ given in the statement of the Lemma. To check that this is a semiring, consider any two sets A and B of this form. We must show that $A \smallsetminus B$ is a finite union of elements of $\mathcal{A}$. We may write A as $\bigcap_{i=1}^{n}[x_i \to E_i]$, B as $\bigcap_{j=1}^{m}[x'_j \to E'_j]$. Then $A \smallsetminus B$ is the finite union of the sets A_j, $1 \le j \le m$, defined by: if $x'_j = x_k$ for some (unique) k, $1 \le k \le n$, then $A_j = \bigcap_{\substack{i=1 \\ i \ne j}}^{n}[x_i \to E_i] \cap [x_k \to E_i \smallsetminus E'_j]$, else $A_j = \bigcap_{i=1}^{n}[x_i \to E_i] \cap [x'_j \to \Omega \smallsetminus E'_j]$.

Note that $\mathcal{A}$ generates the weak σ-algebra on $Sel(F)$, by definition. Assume there is a map $\mu \colon \mathcal{A} \to [0, +\infty]$ that satisfies the formula given in the statement of the Lemma: $\mu(\bigcap_{i=1}^{n}[x_i \to E_i]) = \prod_{i=1}^{n} g(x_i)(E_i)$. For now, this is an assumption, not a definition. For it to be a definition, we would need to check that this is unambiguous: if $\bigcap_{i=1}^{n}[x_i \to E_i] = \bigcap_{j=1}^{m}[x'_j \to E'_j]$, we should verify that $\prod_{i=1}^{n} g(x_i)(E_i) = \prod_{j=1}^{m}[x'_j \to E'_j]$. This will be easier to prove later. Until then, we concentrate on the more interesting question of σ-additivity.

Let $A_k = \bigcap_{i \in I_k}[x_{ki} \to E_{ki}]$, $k \in \mathbb{N}$, be a countable family of disjoint elements of $\mathcal{A}$ whose union is some element A of $\mathcal{A}$ again, where each index set I_k is finite. We must show that $\mu(A) = \sum_{k=0}^{+\infty} \mu(A_k)$.

There is a simple trick to prove this: we exhibit a measure ($s_J[\mu_J]$ below) that coincides with μ on A and each A_k, $k \in \mathbb{N}$. We shall call this the *Łomnick-Ulam trick*, since this also subtends the classical proof of the Łomnick-Ulam theorem.

Observe that the set of points $(x_{ki})_{k \in \mathbb{N}, i \in I_k}$ is countable. For each countable set J of points of X, let $s_J \colon \prod_{x \in J} F(x) \to Sel(F)$ be the map $\vec{\mathbf{y}} \mapsto f_{\vec{\mathbf{y}}}$ given in Lemma 5. Let μ_J be the product probability measure $\bigotimes_{x \in J} g(x)$ on $\prod_{x \in J} F(x)$, as given in Lemma 3.

By definition, $s_J[\mu_J](\mathcal{E}) = \mu_J(s_J^{-1}(\mathcal{E}))$ for every measurable subset $\mathcal{E}$ of $Sel(F)$. In particular, if J contains all the points $x_1, \ldots, x_n$, then $s_J[\mu_J](\bigcap_{i=1}^{n}[x_i \to E_i]) = \mu_J\{\vec{\mathbf{y}} \in \prod_{x \in J} F(x) \mid \forall i, 1 \le i \le n \cdot y_{x_i} \in E_i\}$ (since $s_J(\vec{\mathbf{y}})(x_i) = f_{\vec{\mathbf{y}}}(x_i) = y_{x_i}$, where we agree to write the tuple $\vec{\mathbf{y}}$ with indices in J, namely, as $(y_x)_{x \in J}) = \mu_J(\bigcap_{i=1}^{n} \pi_{x_i}^{-1}(E_i))$. By definition of μ_J, this is equal to $\prod_{i=1}^{n} g(x_i)(E_i)$, hence to $\mu(\bigcap_{i=1}^{n}[x_i \to E_i])$.

In other words, $s_J[\mu_J]$ coincides with μ on all subsets of $\mathcal{A}$ of the form $\bigcap_{i=1}^{n}[x_i \to E_i]$ where every x_i is in J. It is certainly not the case in general that $s_J[\mu_J]$ and μ coincide! The condition that every x_i is in J is crucial. This condition is satisfied by A and every A_k, $k \in \mathbb{N}$, provided we take $J = (x_{ki})_{k \in \mathbb{N}, i \in I_k}$. Since $s_J[\mu_J]$ is σ-additive, the equation $\mu(A) = \sum_{k=0}^{+\infty} \mu(A_k)$ holds.

This construction also shows that μ indeed exists, something we had deferred the verification of. The problem was to show that if $\bigcap_{i=1}^{n}[x_i \to E_i] = \bigcap_{j=1}^{m}[x'_j \to E'_j]$ then $\prod_{i=1}^{n} g(x_i)(E_i) = \prod_{j=1}^{m}[x'_j \to E'_j]$. Take $J = \{x_1, \cdots, x_n\} \cup \{x'_1, \cdots, x'_m\}$. The inverse image by s_J of the set $\bigcap_{i=1}^{n}[x_i \to E_i] = \bigcap_{j=1}^{m}[x'_j \to$

$E'_j]$ is equal to $\bigcap_{i=1}^{n} \pi_{x_i}^{-1}(E_i) = \bigcap_{j=1}^{m} \pi_{x'_j}^{-1}(E'_j)$, and its μ_J-measure is $\prod_{i=1}^{n} g(x_i)$ $(E_i) = \prod_{j=1}^{m}[x'_j \to E'_j]$, by definition of the product probability measure.

The existence of ϖ_g follows directly from Carathéodory's theorem. It is clear that it is a probability measure. Uniqueness follows from the fact that probability measures are uniquely determined by their values on any π-system that generates the σ-algebra [2, Theorem 3.3]. A π-*system* is a collection of sets that is closed under binary intersections, and certainly $\mathcal{A}$ qualifies, as a semiring. □

4 IT Automata

Consider the following simple form of non-deterministic automaton, which we call an *IT automaton* (for Ionescu-Tulcea automaton): a tuple (Z, Λ, F) where Z, Λ are measurable spaces, and $F\colon Z\Lambda^* \to \mathbb{P}^*(\Lambda)$ is a measurable multifunction with a measurable selection. Z can be thought as a space of input values, Λ as a space of states, and F as a generalized transition relation, which given a finite history $z\lambda_1 \cdots \lambda_n$ produces a non-empty set $F(z\lambda_1 \cdots \lambda_n)$ of possible next states. The idea is that the system starts with some input value z, goes to the first state $\lambda_1 \in F(z)$, then to a second state $\lambda_2 \in F(z\lambda_1)$, ..., to an nth state $\lambda_n \in F(z\lambda_1 \cdots \lambda_{n-1})$, and so on. In other words, IT automata are just like non-deterministic automata, except on possibly infinite state spaces and with a history-dependent transition relation.

We use the notation Λ^* for the space of all finite words on the alphabet Λ, which we equate with the countable coproduct $\sum_{n\in\mathbb{N}} \Lambda^n$. We also write $Z\Lambda^*$ instead of $Z \times \Lambda^*$, and will more generally drop the $\times$ symbol in cases where this is not ambiguous. Accordingly, we write $z\lambda_1 \cdots \lambda_n$ in word notation, instead of as the tuple $(z, \lambda_1, \cdots, \lambda_n)$. We have already done so before.

Since an IT automaton starting from input value z will produce infinitely many states $\lambda_1, \ldots, \lambda_n$, it is natural to study the space $Z\Lambda^{\mathbb{N}}$ of *infinite paths* of the automaton, where $\Lambda^{\mathbb{N}}$ is the product of countably infinitely many copies of Λ. (This is written Λ^ω in language theory.) The σ-algebra on $Z\Lambda^{\mathbb{N}}$ is generated by so-called *cylinders*, which are exactly the products $X_0 \Lambda_1 \Lambda_2 \ldots \Lambda_n \Lambda^{\mathbb{N}}$ with $X_0 \in \Sigma_Z$ and $\Lambda_i \in \Sigma_\Lambda$, $1 \le i \le n$, and $n \in \mathbb{N}$.

A *randomized scheduler* for such an IT automaton is a measurable map $g\colon Z\Lambda^* \to \mathcal{P}(\Lambda)$ such that $g(z\lambda_1 \cdots \lambda_n)$ is concentrated on $F(z\lambda_1 \cdots \lambda_n)$ for every $z\lambda_1 \cdots \lambda_n \in Z\Lambda^*$.

Given a measurable subset $\mathcal{E}$ of $Z\Lambda^{\mathbb{N}}$, and an input value $z \in |Z|$, the probability that the induced infinite path lies in $\mathcal{E}$, where λ_n is chosen at random with probability $g(z\lambda_1 \cdots \lambda_{n-1})$ at each step, is given by Theorem 2, with $Q_0 = Z\Lambda^*$, $Q_i = \Lambda$ for every $i \ge 1$, and letting g_n be the restriction of g to $Z\Lambda^{n-1}$. Explicitly:

Proposition 2. *Let Z, Λ be two measurable spaces, and $g\colon Z\Lambda^* \to \mathcal{P}(\Lambda)$ be a measurable map. There is a unique map $P_g\colon Z\Lambda^* \to \mathcal{P}(Z\Lambda^{\mathbb{N}})$ such that $P_g(w)(X_0$ $\Lambda_1 \Lambda_2 \cdots \Lambda_n \cdots \Lambda_{n+m}\Lambda^{\mathbb{N}})$ is equal to:*

$$\chi_{X_0}(z)\chi_{\Lambda_1}(\lambda_1) \cdots \chi_{\Lambda_n}(\lambda_n) \int_{\lambda_{n+1}\in\Lambda_{n+1}} dg(w) \int_{\lambda_{n+2}\in\Lambda_{n+2}} dg(w\lambda_{n+1}) \cdots \int_{\lambda_{n+m}\in\Lambda_{n+m}} dg(w\lambda_{n+1}\cdots\lambda_{n+m-1})$$

for all measurable subsets X_0 of Z, Λ_i of Λ ($1 \le i \le m$, $n \le m$), and elements $w = z\lambda_1 \cdots \lambda_n \in Z\Lambda^*$. Moreover, P_g is measurable.

A *pure scheduler* for the IT automaton (Z, Λ, F) is just a measurable selection of F: given the history $z\lambda_1 \cdots \lambda_{n-1}$, pick a next state λ_n from $F(z\lambda_1 \cdots \lambda_{n-1})$.

Given a pure scheduler f, define $\widetilde{f} \colon Z\Lambda^* \to Z\Lambda^{\mathbb{N}}$ so that $\widetilde{f}(z\lambda_1 \cdots \lambda_n)$ is the unique infinite path that we obtain by starting with the history $z\lambda_1 \cdots \lambda_n$ and repeatedly computing next states, using f: $\widetilde{f}(z\lambda_1 \cdots \lambda_n) = z\lambda_1 \cdots \lambda_n \lambda_{n+1} \cdots \lambda_{n+m} \cdots$ where $\lambda_{n+k+1} = f(z\lambda_1 \cdots \lambda_n \lambda_{n+1} \cdots \lambda_{n+k})$, for every $k \in \mathbb{N}$.

One might think of doing the following. Fix an IT automaton (Z, Λ, F), and a randomized scheduler g for this automaton. Pick a pure scheduler f at random, with respect to the probability measure ϖ_g given by Theorem 3, and show that the probability that $\widetilde{f}(z)$ falls into any given measurable set $\mathcal{E}$ of infinite paths is equal to the probability $P_g(z)(\mathcal{E})$ given in Proposition 2. We did the computation, and checked that this indeed holds... except this is all wrong! This does not make sense unless the map $f \mapsto \widetilde{f}(z)$ (for $z \in |Z|$ fixed) is measurable. We cannot dismiss the problem: this is *the* central question here.

To state it another way, the weak σ-algebra on the space of pure schedulers has to be replaced by a larger one: the $\widetilde{weak\ \sigma\text{-}algebra}$ on the set $|Sel(F)|$ of pure schedulers is the smallest that makes every map $f \in |Sel(F)| \mapsto \widetilde{f}(z\lambda_1 \cdots \lambda_n)$ measurable, for every $z\lambda_1 \cdots \lambda_n \in Z\Lambda^*$. We write $\widetilde{Sel}(F)$ for $|Sel(F)|$ with the $\widetilde{weak}$ σ-algebra.

For $w = z\lambda_1 \cdots \lambda_n \in Z\Lambda^*$, and a measurable subset $\mathcal{E}$ of $\Lambda^{\mathbb{N}}$, let us write $w\mathcal{E}$ for $\{z\} \times \{\lambda_1\} \times \cdots \times \{\lambda_n\} \times \mathcal{E}$. The $\widetilde{weak}$ σ-algebra is generated by the sets $[w \overset{\sim}{\to} w\mathcal{E}]$, defined as the set of pure schedulers f such that $\widetilde{f}(w) \in w\mathcal{E}$.

We wish to define our probability measure $\widetilde{\omega}_g$ on pure schedulers f by saying that the probability that $\widetilde{f}(w) \in w\mathcal{E}$ (for any fixed w, $\mathcal{E}$) is exactly $P_g(w)(\mathcal{E})$, where P_g is given in Proposition 2: namely, $\widetilde{\omega}_g([w \overset{\sim}{\to} w\mathcal{E}]) = P_g(w)(\mathcal{E})$. That cannot be enough to define $\widetilde{\omega}_g$, and we need to at least define $\widetilde{\omega}_g(\bigcap_{i=1}^n [w_i \overset{\sim}{\to} w_i\mathcal{E}_i])$ for all finite intersections of sets $[w_i \overset{\sim}{\to} w_i\mathcal{E}_i]$. Now there is a big difference with the case of random measurable selections (Section 3): the choices we make for $\widetilde{f}(w_i)$ for different indices i *cannot in general be independent*. Indeed, imagine we have chosen $\widetilde{f}(w_i)$, for some i, to be $w_i\lambda_{n+1} \cdots \lambda_{n+m} \cdots$: then we have no choice for $\widetilde{f}(w_i\lambda_{n+1})$, and also for $\widetilde{f}(w_i\lambda_{n+1}\lambda_{n+2})$, ..., which must all be equal to $w_i\lambda_{n+1} \cdots \lambda_{n+m} \cdots$. This is the *consistency* condition mentioned in the proof of Proposition 3 below: in general, if we have chosen $\widetilde{f}(w_i)$ as $w_i\omega_i$ for some $\omega_i \in \Lambda^{\mathbb{N}}$, and later we need to choose $\widetilde{f}(w_j)$ where w_i is a prefix of w_j, and w_j is a prefix of $w_i\omega_i$, then we must choose it as $\widetilde{f}(w_j) = w_i\omega_i$.

We still proceed in a manner similar to Section 3. We now need an extra assumption: say that measurable space X *has measurable diagonal* if and only if the diagonal $\Delta = \mathrm{Gr}(=) = \{(x, x) \mid x \in |X|\}$ is measurable in $X \times X$. Dravecký [9, Theorem 1] shows a number of equivalent conditions. One of these is that X has measurable diagonal if and only if there is a countable family $(E_n)_{n \in \mathbb{N}}$ of measurable subsets of X that separates points, that is, such that for any two

distinct points x, y, there is an E_n that contains one and not the other. This is true for all Polish spaces, notably. Another one is that X has measurable diagonal if and only if for every measurable function f from an arbitrary measurable space Y to X, the obvious multifunction $y \in Y \mapsto \{f(y)\}$ is measurable. This apparent tautology is wrong when X does not have measurable diagonal! The canonical counter-example is $Y = X$, $f = \mathrm{id}_X$: the corresponding multifunction is precisely the one whose graph is Δ. Finally, every one-element subset $\{x\}$ of a space X with measurable diagonal is measurable in X; indeed, $\{x\}$ is the vertical cut $\Delta_{|x}$.

Proposition 3. *Let (Z, Λ, F) be an IT automaton, with a randomized scheduler $g \colon Z\Lambda^* \to \mathcal{P}(\Lambda)$. Assume that Z and Λ both have measurable diagonals. Let $P_g \colon Z\Lambda^* \to \mathcal{P}(Z\Lambda^{\mathbb{N}})$ be the probability-on-paths map given in Proposition 2. There is a probability measure $\widetilde{\omega}_g$ on $\widetilde{Sel}(F)$ such that $\widetilde{\omega}_g([z \overset{\sim}{\to} \mathcal{E}]) = P_g(z)(\mathcal{E})$ for all $z \in Z$ and $\mathcal{E} \in \Sigma_{Z\Lambda^{\mathbb{N}}}$.*

Proof. (Outline. The technical details are relegated to Appendix A.) As a notational help, we write w, possibly subscripted or primed, for finite words in $Z\Lambda^*$, and ω, possibly subscripted or primed, for infinite words in $\Lambda^{\mathbb{N}}$.

Let us write $\preceq$ for the prefix relation on finite and infinite words. Say that a set of words A of $Z\Lambda^*$ is *prefix-closed* if and only if for every $w \in Z\Lambda^*$ that is a prefix of some element of A, w is in A as well.

Let $\mathcal{A}$ be the semiring of all finite intersections of basic measurable subsets $\bigcap_{i=1}^{n}[w_i \overset{\sim}{\to} w_i\mathcal{E}_i]$. By adding extra words if needed, we may assume that $\mathcal{I} = \{w_1, \ldots, w_n\}$ is prefix-closed.

It is hard to even attempt to describe explicitly the values of $\widetilde{\omega}_g$ on elements of $\mathcal{A}$. In the proof of Theorem 3, we had eventually shown that the value of the desired measure ϖ_g coincided with the image measure of some other measure defined on a product space for sufficiently small parts of the semiring. We define $\widetilde{\omega}_g$ on $\mathcal{A}(\mathcal{I})$ by a similar Łomnick-Ulam-like trick: through image measures of some measures $\mu_{\mathcal{W}}$ under maps $\alpha_{\mathcal{W}}$, for countably infinite subsets $\mathcal{W}$ of $Z\Lambda^*$.

Given a countably infinite set $\mathcal{W} = (w_j)_{j \in J}$ of words in $Z\Lambda^*$ (with $J = \mathbb{N} \smallsetminus \{0\}$, say), let $Cst_{\mathcal{W}}$ be the subspace of $\prod_{w \in \mathcal{W}} Path_F(w)$ consisting of those tuples $(w_j\omega_j)_{j \in J}$ that are *consistent*: for all $i, j \in J$ such that $w_i \preceq w_j \preceq w_i\omega_i$, then $w_i\omega_i = w_j\omega_j$. In pictures, if the leftmost two zones are equal then the rightmost zones are equal, too:

Using Ionescu-Tulcea's Theorem 2, we build a probability measure $\mu_{\mathcal{W}}$ on $Cst_{\mathcal{W}}$ (Proposition 5 in Appendix A). Intuitively, $\mu_{\mathcal{W}}$ picks $w_1\omega_1$ at random using probability measure $P_g(w_1)$, then picks $w_2\omega_2$, ..., $w_j\omega_j$, ..., as follows. At step j, if $w_i \preceq w_j \preceq w_i\omega_i$ for some previous i, $1 \leq i < j$, then we pick $w_j\omega_j$, deterministically, as equal to $w_i\omega_i$, enforcing consistency; otherwise, we pick $w_j\omega_j$ at random using probability measure $P_g(w_j)$. All this makes sense provided

we sort $\mathcal{W}$ topologically, i.e., we choose the indexing scheme so that any w_i that is a prefix of w_j occurs before w_j, viz., $i \leq j$ (Lemma 6 in Appendix A).

Given $w \in \mathcal{W}$, say $w = w_j$, let us write $[w : \mathcal{E}]$ for the set of consistent tuples $\overrightarrow{\mathbf{w}\omega}$ in $Cst_{\mathcal{W}}$ such that $\omega_j \in \mathcal{E}$. One can show that, given any finite prefix-closed set $\{w_1, \ldots, w_n\}$ of words in $Z\Lambda^*$, the value $\mu_{\mathcal{W}}(\bigcap_{i=1}^{n}[w_i : \mathcal{E}_i])$ is the same for *all* countably infinite subsets $\mathcal{W}$ that contain $w_1, \ldots, w_n$ (Lemma 9 in Appendix A). The formula we obtain simplifies when $n = 1$ and w_1 is a single letter $z \in Z$: $\mu_{\mathcal{W}}([z : \mathcal{E}]) = P_g(z)(z\mathcal{E})$ (Lemma 10 in Appendix A).

We take the image measure of $\mu_{\mathcal{W}}$ under a measurable map $\alpha_{\mathcal{W}} \colon \overrightarrow{\mathbf{w}\omega} \in Cst_{\mathcal{W}} \mapsto f_{\overrightarrow{\mathbf{w}\omega}} \in \widetilde{Sel}(F)$ that retrieves a canonical pure scheduler from a consistent set of tuples (Proposition 6 in Appendix A), defined in such a way that $\widetilde{f}_{\overrightarrow{\mathbf{w}\omega}}(w_j) = w_j\omega_j$ for every $j \in J$. This is done by patching, similarly to Lemma 5.

Note that $\alpha_{\mathcal{W}}[\mu_{\mathcal{W}}](\bigcap_{i=1}^{n}[w_i \overset{\sim}{\to} w_i\mathcal{E}_i])$ is independent of $\mathcal{W}$, provided that $\mathcal{W}$ contains the prefix-closed subset $\{w_1, \ldots, w_n\}$. Indeed, $\alpha_{\mathcal{W}}[\mu_{\mathcal{W}}](\bigcap_{i=1}^{n}[w_i \overset{\sim}{\to} w_i\mathcal{E}_i]) = \mu_{\mathcal{W}}(\bigcap_{i=1}^{n}[w_i : \mathcal{E}_i])$, which we have shown independent of $\mathcal{W}$. We can therefore define $\widetilde{\omega}_g$ as coinciding with $\alpha_{\mathcal{W}}[\mu_{\mathcal{W}}]$ on those elements $\bigcap_{i=1}^{n}[w_i \overset{\sim}{\to} w_i\mathcal{E}_i]$ of $\mathcal{A}$ with $\{w_1, \ldots, w_n\} \subseteq \mathcal{W}$. As such, it is σ-additive on $\mathcal{A}$: as in the proof of Theorem 3, let $A_k = \bigcap_{w \in \mathcal{I}_k}[w \overset{\sim}{\to} \mathcal{E}_{wi}]$, $k \in \mathbb{N}$, be a countable family of disjoint elements of $\mathcal{A}$, where $\mathcal{I}_k$ is finite and prefix-closed, and $A = \bigcap_{w \in \mathcal{I}}[w \overset{\sim}{\to} \mathcal{E}_w]$ be their union, assumed in $\mathcal{A}$, with $\mathcal{I}$ prefix-closed again. Then $\widetilde{\omega}_g(A) = \sum_{k \in \mathbb{N}} \widetilde{\omega}_g(A_k)$, since $\widetilde{\omega}_g$ coincides with the measure $\alpha_{\mathcal{W}}[\mu_{\mathcal{W}}]$ on A and every A_k, for $\mathcal{W} = \bigcup_{k \in \mathbb{N}} \mathcal{I}_k \cup \mathcal{I}$ (which is countable).

Finally, $\widetilde{\omega}_g([z \overset{\sim}{\to} z\mathcal{E}]) = \mu_{\mathcal{W}}([z : \mathcal{E}]) = P_g(z)(z\mathcal{E})$, for every measurable subset $\mathcal{E}$ of $\Lambda^{\mathbb{N}}$. Since $P_g(z)$ is concentrated on $Path_F(z)$, hence on $z\Lambda^{\mathbb{N}}$, for every measurable subset $\mathcal{E}$ of $Z\Lambda^{\mathbb{N}}$, $\widetilde{\omega}_g([z \overset{\sim}{\to} \mathcal{E}]) = \widetilde{\omega}_g([z \overset{\sim}{\to} z\mathcal{E}_{|z}]) = P_g(z)(z\mathcal{E}_{|z}) = P_g(z)(\mathcal{E} \cap z\Lambda^{\mathbb{N}}) = P_g(z)(\mathcal{E})$. $\qquad\square$

We can now integrate on infinite paths ω with respect to $P_g(z)$, or on pure schedulers, and this will give the same average value:

Proposition 4. *Under the assumptions of Proposition 3, let $h \colon Z\Lambda^{\mathbb{N}} \to \mathbb{R}^+$ be a measurable map, and $z \in Z$, then:*

$$\int_{f \in \widetilde{Sel}(F)} h(\widetilde{f}(z))d\widetilde{\omega}_g = \int_{\omega \in Z\Lambda^{\mathbb{N}}} h(\omega)dP_g(z).$$

Proof. Let $\widetilde{\alpha}_z \colon f \mapsto \widetilde{f}(z)$. This is a measurable map, since $\widetilde{\alpha}_z^{-1}(\mathcal{E}) = [z \overset{\sim}{\to} \mathcal{E}]$. The left-hand side is $\int_{f \in \widetilde{Sel}(F)} h(\widetilde{\alpha}_z(f))d\widetilde{\omega}_g = \int_{\omega \in Z\Lambda^{\mathbb{N}}} h(\omega)d\widetilde{\alpha}_z[\widetilde{\omega}_g]$ by the change of variables formula (1). Proposition 3 states precisely that $\widetilde{\alpha}_z[\widetilde{\omega}_g] = P_g(z)$. $\qquad\square$

We may think of h as a *payoff* function on infinite paths. The above shows that the average payoff with respect to $P_g(z)$ is also the average of the individual payoffs $h(\widetilde{f}(z))$ one would get by drawing a pure scheduler f at random instead.

Fact 1 then implies that the value of the average payoff is bounded by the payoff evaluated on two pure schedulers f^- and f^+:

Corollary 1 (Randomness for Free). *Under the assumptions of Proposition 3, let $h\colon Z\Lambda^{\mathbb{N}} \to \mathbb{R}^+$ be a measurable map, and $z \in Z$. There are two pure schedulers f^- and f^+ in $Sel(F)$ such that:*

$$h(\widetilde{f^-}(z)) \leq \int_{\omega \in Z\Lambda^{\mathbb{N}}} h(\omega)dP_g(z) \leq h(\widetilde{f^+}(z)).$$

5 Conclusion

We have established a few "randomness for free" type theorems for measurable multifunctions first, for IT automata second. The results are pleasing, and our assumptions are fairly minimal. Our proofs use fairly simple ideas, too: there is the Chatterjee-Doyen-Gimbert-Henzinger idea of drawing measurable selections/pure schedulers at random first, and this makes sense because of a combination of patching, of Carathéodory's measure extension theorem, and of a Łomnik-Ulam type trick.

Where should we go next? One may push the results on IT automata to *partially observable IT automata*. Instead of a measurable transition multifunction $F\colon Z\Lambda^* \to \mathbb{P}^*(\Lambda)$, such automata have a measurable transition multifunction $F\colon Z\Lambda^* \to \mathbb{P}^*(\Xi)$, where Ξ is an (additional) measurable space Ξ of *actual states*. Such states q are mapped to *observable* states $\lambda \in \Lambda$ by a measurable map $\tau\colon Z\Lambda^* \times \Xi \to \Lambda$; this may depend on the past history $w \in Z\Lambda^*$, viz., $\lambda = \tau(w, q)$. We pick the next observable state λ after history w by picking q from $F(w)$, then computing $\tau(w, q)$. Modifying the notion of consistent paths as required, it seems feasible to prove an analogue of Proposition 3 for partially observable IT automata. The σ-algebra on $Sel(F)$ needs to be changed again! so that we cannot reuse Proposition 3 as is. Once this is done, we can proceed to stochastic automata [4]. Given a stochastic automaton with state space Q and action space L, take $Z = Q$, $\Lambda = (L \times Q)_\perp$ (writing $X_\perp$ for $X + \{\perp\}$), and $\Xi = \mathcal{P}(\Lambda)$. Taking g to be the second projection map π_2, Theorem 3 allows us to draw the observation maps $\tau\colon Q(L \times Q)^*_\perp \times \mathcal{P}((L \times Q)_\perp) \to (L \times Q)_\perp$ at random, and this will simulate the probabilistic choice of q' with respect to μ described in the introduction. Combining this with the alluded "randomness for free" result for partially observable IT automata, we hope that it would settle the "randomness for free" question for general, stochastic automata.

References

1. Assalé Adjé, Olivier Bouissou, Jean Goubault-Larrecq, Éric Goubault, and Sylvie Putot. Analyzing numerical programs with partially known distributions. In Ernie Cohen and Andrey Rybalchenko, editors, *Proc. 5th IFIP TC2 WG 2.3 Conf. Verified Software—Theories, Tools, and Experiments (VSTTE'13)*, Lecture Notes in Computer Science. Springer Verlag, 2013. `http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/ABGGP-vstte13.pdf`.
2. Patrick Billingsley. *Probability and Measure*. Wiley series in probability and mathematical statistics. John Wiley and Sons, 3rd edition, 1995.

3. David Harold Blackwell. A Borel set not containing a graph. *Annals of Mathematical Statistics*, 39:1345–1347, 1968.

4. Stefano Cattani, Roberto Segala, Marta Kwiatkowska, and Gethin Norman. Stochastic transition systems for continuous state spaces and non-determinism. In *Proc. 8th Intl. Conf. Foundations of Software Science and Computational Structures (FOSSACS'05)*, pages 125–139. Springer Verlag LNCS 3441, 2005.

5. Krishnendu Chatterjee, Laurent Doyen, Hugo Gimbert, and Thomas A. Henzinger. Randomness for free. In *Proc. 35th Intl. Symp. Mathematical Foundations of Computer Science (MFCS'10)*, pages 246–257. Springer Verlag LNCS 6281, 2010.

6. Josée Desharnais, Abbas Edalat, and Prakash Panangaden. Bisimulation for labelled Markov processes. *Information and Computation*, 179(2):163–193, 2002.

7. Josée Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Metrics for labelled Markov processes. *Theoretical Computer Science*, 318:323–354, 2004.

8. Ernst-Erich Doberkat. *Stochastic Relations: Foundations for Markov Transition Systems*. Studies in Informatics Series. Chapman and Hall/CRC, 2007.

9. Jozef Dravecký. Spaces with measurable diagonal. *Matematický Časopis*, 25(1):3–9, 1975.

10. Norm Ferns, Prakash Panangaden, and Doina Precup. Bisimulation metrics for continuous Markov decision processes. *SIAM Journal on Computing*, 40(6):1662–1714, 2011.

11. Michèle Giry. A categorical approach to probability theory. In Bernhard Banaschewski, editor, *Categorical Aspects of Topology and Analysis*, pages 68–85. Springer Verlag LNM 915, 1981.

12. Charles J. Himmelberg. Measurable relations. *Fundamenta Mathematicae*, 87:53–71, 1975.

13. Olav Kallenberg. *Foundations of Modern Probability*. Probability and its Applications. Springer, 2nd edition edition, 2002.

14. Prakash Panangaden. Probabilistic relations. In C. Baier, M. Huth, M. Kwiatkowska, and M. Ryan, editors, *Proceedings of PROBMIV'98*, pages 59–74, 1998. `http://www.cs.bham.ac.uk/~mzk/probmiv/prelproc98/`.

15. Prakash Panangaden. Probability and measure for concurrency theorists. *Theoretical Computer Science*, 253(2):287–309, February 2001.

16. Prakash Panangaden. *Labelled Markov Processes*. Imperial College Press, 2009.

17. Sashi Mohan Srivastava. *A Course on Borel Sets*, volume 180 of *Graduate Texts in Mathematics*. Springer Verlag, 1998.

18. Daniel H. Wagner. Survey of measurable selection theorems: An update. In *Measure Theory, Oberwolfach, 1979*, pages 176–219. Springer Verlag LNM 794, 1980.

A Auxiliary Results Needed for Proposition 3

In the following, we fix an IT automaton (Z, Λ, F) with a randomized scheduler $g \colon Z\Lambda^* \to \mathcal{P}(\Lambda)$. We also assume that Z and Λ both have measurable diagonals. Let $P_g \colon Z\Lambda^* \to \mathcal{P}(Z\Lambda^{\mathbb{N}})$ be the probability-on-paths map given in Proposition 2.

In the rest of this section, we fix a countably infinite set $\mathcal{W}$ of words in $Z\Lambda^*$. We let $J = \{1, 2, \cdots\} = \mathbb{N} \smallsetminus \{0\}$. This will serve as an index set.

Lemma 6. *One can write $\mathcal{W}$ as a family $(w_j)_{j \in J}$ in such a way that for all $i, j \in J$, if $w_i \preceq w_j$ then $i \leq j$.*

In other words, one can sort the words topologically. Proving this involves showing that the order type of $\mathcal{W}$ under the prefix ordering is at most ω.

Proof. There are only countably many elements of Z, resp. Λ, that can occur in any word from $\mathcal{W}$, so one can attribute each of them a unique natural number. Equate each element with the corresponding natural number. Each word $w = z\lambda_1 \cdots \lambda_n$ can now be encoded as $p_1^z p_2^{\lambda_1} \cdots p_{n+1}^{\lambda_n}$, where p_1, p_2, $\ldots$, p_n, $\ldots$ enumerate the prime numbers. This way, if w_i is a proper prefix of w_j, then w_i will be encoded as a lower number than w_j. Now enumerate the words w_i in the order of their encodings. $\square$

In the rest of Section A, we assume such a topologically sorted indexing scheme $(w_j)_{j \in J}$ for $\mathcal{W}$.

Lemma 7. *For every $n \in J$, $P_g(w_n)$ is concentrated on $Path_F(w_n)$.*

Proof. Since Z and Λ have measurable diagonals, so does $Z\Lambda^*$. For each $n \in J$, $\{w_n\}$ is therefore measurable in $Z\Lambda^*$, so we can apply Lemma 2 and conclude that $P_g(w_n)$ is concentrated on $Path_F(w_n)$. $\square$

In particular, the restriction $P_g(w_n)_{|Path_F(w_n)}$ makes sense. To reduce visual clutter, we simply write $P_g(w_n)$ for the latter probability measure on $Path_F(w_n)$.

We shall write $w_n\Lambda^{\mathbb{N}}$ for the set of words in $Z\Lambda^{\mathbb{N}}$ that have w_n as a prefix. As the product $\{w_n\} \times \Lambda^{\mathbb{N}}$, this is a measurable set. (Recall that one-element sets are measurable, since Z and Λ have measurable diagonals.)

We wish to draw the values $w_j\omega_j$ of $\widetilde{f}(w_j)$, $j \in J$, in a consistent way, namely, if we have already mapped w_i to the value $w_i\omega_i$, and $w_i \preceq w_j \preceq w_i\omega_i$, then we have no choice and must choose to map w_j to $w_i\omega_i$ as well. We achieve this by using Ionescu-Tulcea's Theorem 2 on *another* probabilistic transition system, defined as follows, and which we only use as a mathematical helper. This is certainly no real-life, practical transition system, and is not meant to be.

We let $Q_0 = \{*\}$, $q_0 = *$, $Q_n = Path_F(w_n)$ for $n \in J$, and $g_n \colon \prod_{i=0}^{n-1} Q_i \to \mathcal{P}(Q_n)$ map $(*, w_1\omega_1, \cdots, w_{n-1}\omega_{n-1})$ to:

- the Dirac mass $\delta_{w_i\omega_i}$, where i is the least index in $1, \cdots, n-1$ such that $w_i \preceq w_n \preceq w_i\omega_i$, if one such index exists; this implements consistency;
- otherwise, the probability measure $P_g(w_n)$.

Both are probability measures on $Q_n = Path_F(w_n)$: the second one by Lemma 7, the first one because $w_i\omega_i$ is a path in $Path_F$ starting with w_n.

Lemma 8. *For every $n \in J$, g_n is measurable.*

Proof. For every i, $1 \le i < n$, such that $w_i \preceq w_n$, the set A_i of infinite words $w_i\omega_i$ in $Path_F(w_i)$ that have w_n as a prefix is $Path_F(w_i) \cap w_n\Lambda^{\mathbb{N}}$, hence is measurable. By extension, when $w_i \not\preceq w_n$, write A_i for the empty set. Rephrasing the definition, $g_n(*, w_1\omega_1, \cdots, w_{n-1}\omega_{n-1})$ is defined as $\mathfrak{e}_{Path_F(w_n)}(w_i\omega_i)$ on $E_i = \{*\} \times \overline{A_1} \times \cdots \times \overline{A_{i-1}} \times A_i \times |Q_{i+1}| \times \cdots \times |Q_{n-1}|$, $1 \le i < n$ (recall that $\mathfrak{e}$

is the monad unit, which is a measurable map), and as $P_g(w_n)$ on $\overline{\bigcup_{i=1}^{n-1} E_i}$ (a constant map). So g_n is a patch of measurable maps, and is therefore measurable by Lemma 4. $\qquad\square$

We can now apply Ionescu-Tulcea's Theorem, as promised, and obtain a probability measure on $\{*\} \times \prod_{j \in J} Path_F(w_j) \cong \prod_{j=1}^{+\infty} Path_F(w_j)$ that we decide to call $\mu_\mathcal{W}$. (This is the measure $P_{g_*}(*)$ of Theorem 2, but we wish to avoid any visual confusion with P_g.)

Let $Cst_\mathcal{W}$ be the set of tuples of paths $(w_j\omega_j)_{j \in J}$ in $\prod_{j \in J} Path_F(w_j)$ that are *consistent*, i.e., such that for all $i, j \in J$ such that $w_i \preceq w_j \preceq w_i\omega_i$, $w_j\omega_j = w_i\omega_i$. These are the only ones we can ever hope to produce from a pure scheduler f, namely, that are of the form $(\widetilde{f}(w_j))_{j \in J}$.

Proposition 5. *The probability measure $\mu_\mathcal{W}$ is concentrated on the set $Cst_\mathcal{W}$ of consistent tuples.*

Proof. Let $\mathcal{F}_n$ be the multifunction from $\prod_{j=1}^{n-1} Path_F(w_j)$ to $Path_F(w_n)$ defined, similarly to g_n, by letting $\mathcal{F}_n(w_1\omega_1, \cdots, w_{n-1}\omega_{n-1})$ be:

- $\{w_i\omega_i\}$, where i is the least index in $1, \cdots, n-1$ such that $w_i \preceq w_n \preceq w_i\omega_i$, if one such index exists;
- otherwise, $Path_F(w_n)$.

One checks easily that $\mathcal{F}_n$ is a measurable multifunction, and that $g_n(w_1\omega_1, \cdots, w_{n-1}\omega_{n-1})$ is concentrated on $\mathcal{F}_n(w_1\omega_1, \cdots, w_{n-1}\omega_{n-1})$. We apply Lemma 2 to $\mathcal{E}_n = \operatorname{Gr} \mathcal{F}_n$ and obtain that $\mu_\mathcal{W}$ is supported on $Path_\mathcal{E}$.

It remains to show that $Path_\mathcal{E} = Cst_\mathcal{W}$. Given any inconsistent tuple $(w_j\omega_j)_{j \in J}$, there must be two indices $i, j \in J$ such that $w_i \preceq w_j \preceq w_i\omega_i$ but $w_j\omega_j \neq w_i\omega_i$. Since $\mathcal{W}$ is topologically sorted, $i \leq j$. Take i minimal. Then $w_j\omega_j$ would be outside $\{w_i\omega_i\} = \mathcal{F}_j(w_1\omega_1, \cdots, w_{j-1}\omega_{j-1}) = (\mathcal{E}_j)_{|(w_1\omega_1, \cdots, w_{j-1}\omega_{j-1})}$, showing that the tuple is not in $Path_\mathcal{E}$. This establishes that $Path_\mathcal{E} \subseteq Cst_\mathcal{W}$. The converse inclusion is obvious. $\qquad\square$

The restriction $\mu_{\mathcal{W}|Cst_\mathcal{W}}$ therefore makes sense. Again, we simply write $\mu_\mathcal{W}$ for this restriction, and consider it as a probability measure on $Cst_\mathcal{W}$.

Given $w \in \mathcal{W}$, say $w = w_j$, let us write $[w : \mathcal{E}]$ for the set of consistent tuples $\overrightarrow{\mathbf{w}\omega}$ in $Cst_\mathcal{W}$ such that $\omega_j \in \mathcal{E}$.

A set A of words in $Z\Lambda^*$ is *prefix-closed* if and only if, for every $w \in Z\Lambda^*$ and $\lambda \in \Lambda$, $w\lambda \in \mathcal{I}$ implies $w \in \mathcal{I}$.

Lemma 9. *Let $\{w'_1, \ldots, w'_n\}$ be a finite prefix-closed set of words in $Z\Lambda^*$. For all measurable subsets $\mathcal{E}_1, \ldots, \mathcal{E}_n$ of $\Lambda^\mathbb{N}$, the value of $\mu_\mathcal{W}(\bigcap_{i=1}^n [w'_i : \mathcal{E}_i])$ is independent of the countably infinite set $\mathcal{W}$, provided it is a superset of $\{w'_1, \ldots, w'_n\}$.*

Proof. Consider any countably infinite superset $\mathcal{W}$ of $\{w'_1, \ldots, w'_n\}$. Write $\mathcal{W}$ as $(w_j)_{j \in J}$, as usual. Let $w'_1 = w_{j_1}, \ldots, w'_n = w_{j_n}$. Up to permutation, we may assume that $j_1 < j_2 < \cdots < j_n$. Then $\bigcap_{i=1}^n [w'_i : \mathcal{E}_i]$ is equal to the intersection of $Cst_\mathcal{W}$ with $\prod_{j=1}^{j_1-1} Path_F(w_j) \times w_{j_1}\mathcal{E}_1 \times \prod_{j=j_1+1}^{j_2-1} Path_F(w_j) \times w_{j_2}\mathcal{E}_2 \times \cdots \times$

$\prod_{j=j_{n-1}+1}^{j_n-1} Path_F(w_j) \times w_{j_n}\mathcal{E}_n \times \prod_{j=j_n+1}^{+\infty} Path_F(w_j)$. We now use formula (2). This requires an abbreviation for all the integrals $\int_{w_j\omega_j\in Z\Lambda^\mathbb{N}} dg_j(*, w_1\omega_1, \cdots, w_{j-1}\omega_{j-1})$ with $j \notin \{j_1, j_2, \cdots, j_n\}$—which will turn to be useless: write $\iint_{k\cdots\ell} \overrightarrow{\mathbf{dg}}$ for the list of symbols

$$\int_{w_k\omega_k\in Path_F(w_k)} dg_k(*, w_1\omega_1, \cdots, w_{k-1}\omega_{k-1}) \int_{w_{k+1}\omega_{k+1}\in Path_F(w_{k+1})} dg_{k+1}(*, w_1\omega_1, \cdots, w_k\omega_k)$$

$$\cdots \int_{w_\ell\omega_\ell\in Path_F(w_\ell)} dg_\ell(*, w_1\omega_1, \cdots, w_{\ell-1}\omega_{\ell-1}).$$

We can now write:

$$\mu_\mathcal{W}(\bigcap_{i=1}^n [w_i' : \mathcal{E}_i]) = \iint_{1\cdots j_1-1} \overrightarrow{\mathbf{dg}} \int_{w_{j_1}\omega_{j_1}\in w_{j_1}\mathcal{E}_1} dg_{j_1}(*, w_1\omega_1, \cdots, w_{j_1-1}\omega_{j_1-1}) \qquad (4)$$

$$\iint_{j_1+1\cdots j_2-1} \overrightarrow{\mathbf{dg}} \int_{w_{j_2}\omega_{j_2}\in w_{j_2}\mathcal{E}_2} dg_{j_2}(*, w_1\omega_1, \cdots, w_{j_2-1}\omega_{j_2-1}) \cdots$$

$$\iint_{j_{n-1}+1\cdots j_n-1} \overrightarrow{\mathbf{dg}} \int_{w_{j_n}\omega_{j_n}\in w_{j_n}\mathcal{E}_n} dg_{j_n}(*, w_1\omega_1, \cdots, w_{j_n-1}\omega_{j_n-1}).$$

Since $\{w_1, \ldots, w_n\}$ is prefix-closed, for each i, $1 \le i \le n$, $g_{j_i}(*, w_1\omega_1, \cdots, w_{j_i-1}\omega_{j_i-1})$ can be written as a function $g_i'(w_{j_1}\omega_{j_1}, \cdots, w_{j_{i-1}}\omega_{j_{i-1}})$ of just those words $w_k\omega_k$ with $k \in \{j_1, \ldots, j_n\}$, $k < j_i$. Explicitly, $g_i'(w_{j_1}\omega_{j_1}, \cdots, w_{j_{i-1}}\omega_{j_{i-1}})$ is $\delta_{w_{j_\ell}\omega_{j_\ell}}$ where ℓ is the least index, $1 \le \ell < i$, such that $w_{j_\ell} \preceq w_{j_i} \preceq w_{j_\ell}\omega_{j_\ell}$ if one exists, and $P_g(w_{j_i})$ otherwise. In (4), the final integral $\int_{w_{j_n}\omega_{j_n}\in w_{j_n}\mathcal{E}_n} dg_{j_n}(*, w_1\omega_1, \cdots, w_{j_n-1}\omega_{j_n-1}) = \int_{w_{j_n}\omega_{j_n}\in w_{j_n}\mathcal{E}_n} dg_n'(w_{j_1}\omega_{j_1}, \cdots, w_{j_{n-1}}\omega_{j_{n-1}})$ is independent of all the formal variables $w_{j_{n-1}+1}\omega_{j_{n-1}+1}$, ..., $w_{j_n-1}\omega_{j_n-1}$ that the integrals $\int_{w_j\omega\in Path_F(w_j)} dg_j(*, w_1\omega_1, \cdots, w_{j-1}\omega_{j-1})$ hidden in $\iint_{j_{n-1}+1\cdots j_n-1} \overrightarrow{\mathbf{dg}}$ introduce. Since all these integrals are with respect to probability measures, they merely contribute a factor of 1. We repeat the process, from right to left in (4), erasing all the notations $\iint_{j_{n-1}+1\cdots j_n-1} \overrightarrow{\mathbf{dg}}$, and obtain:

$$\mu_\mathcal{W}(\bigcap_{i=1}^n [w_i' : \mathcal{E}_i]) = \int_{w_{j_1}\omega_{j_1}\in w_{j_1}\mathcal{E}_1} dg_{j_1}'() \int_{w_{j_2}\omega_{j_2}\in w_{j_2}\mathcal{E}_2} dg_2'(w_{j_1}\omega_{j_1}) \cdots \int_{w_{j_n}\omega_{j_n}\in w_{j_n}\mathcal{E}_n} dg_n'(w_{j_1}\omega_{j_1}, \cdots, w_{j_{n-1}}\omega_{j_{n-1}}).$$

$$(5)$$

It is now evident that $\mu_\mathcal{W}(\bigcap_{i=1}^n [w_i' : \mathcal{E}_i])$ is independent of $\mathcal{W}$. $\quad\square$

Applying (5) to the case $n = 1$, $w_{j_1} = z \in Z$ (which is automatically prefix-closed), and noticing that $g_{j_1}'() = P_g(w_{j_1}) = P_g(z)$, we obtain:

Lemma 10. *For every $z \in Z$, for every countably infinite set $\mathcal{W}$ of words of $Z\Lambda^*$ containing z, for every measurable subset $\mathcal{E}$ of $\Lambda^\mathbb{N}$, $\mu_\mathcal{W}([z : \mathcal{E}]) = P_g(z)(z\mathcal{E})$.*

Proposition 6. *For every consistent tuple $\overrightarrow{\mathbf{w}\omega} = (w_j\omega_j)_{j\in J}$ in $Cst_\mathcal{W}$, there is a pure scheduler $f_{\overrightarrow{\mathbf{w}\omega}}$ such that $\widetilde{f_{\overrightarrow{\mathbf{w}\omega}}}(w_j) = w_j\omega_j$ for every $j \in J$. Moreover, the map $\alpha_\mathcal{W}\colon \overrightarrow{\mathbf{w}\omega} \in Cst_\mathcal{W} \mapsto f_{\overrightarrow{\mathbf{w}\omega}} \in \widetilde{Sel}(F)$ is measurable.*

Proof. If $w \in Z\Lambda^*$ is a prefix of an infinite word in $\Lambda^{\mathbb{N}}$, then this infinite word can be written in a unique way as $w\lambda\omega$ for some $\lambda \in \Lambda$ and $\omega \in \Lambda^{\mathbb{N}}$: let us call λ the *letter after w* in the infinite word.

Since (Z, Λ, F) is an IT automaton, F has a measurable selection σ. For each consistent tuple $\overrightarrow{\mathbf{w}\omega} = (w_j\omega_j)_{j \in J}$ in $Cst_{\mathcal{W}}$, we define a pure scheduler $f_{\overrightarrow{\mathbf{w}\omega}}$ as follows. For every $w \in Z\Lambda^*$,

- either $w_j \preceq w \preceq w_j\omega_j$ for some $j \in J$, and $f_{\overrightarrow{\mathbf{w}\omega}}(w)$ is the letter after w in $w_j\omega_j$;
- or $w_j \preceq w \preceq w_j\omega_j$ for no $j \in J$, and $f_{\overrightarrow{\mathbf{w}\omega}}(w) = \sigma(w)$.

In the first case, it does not matter which $j \in J$ is picked, because of consistency. Imagine indeed that $w_i \preceq w \preceq w_i\omega_i$ and $w_j \preceq w \preceq w_j\omega_j$ for two indices $i, j \in J$. Since w_i and w_j are two prefixes of the same word w, one of them must be a prefix of the other, say $w_i \preceq w_j$. Then $w_j \preceq w \preceq w_i\omega_i$, and consistency entails that $w_i\omega_i = w_j\omega_j$, so the letter after w is the same in both infinite words.

For short, let us say that w is *stored in $\overrightarrow{\mathbf{w}\omega}$* if and only if $w_j \preceq w \preceq w_j\omega_j$ for some $j \in J$. In this case, it is easy to see that $\widetilde{f_{\overrightarrow{\mathbf{w}\omega}}}(w) = w_j\omega_j$. This implies our first claim, namely that $\widetilde{f_{\overrightarrow{\mathbf{w}\omega}}}(w_j) = w_j\omega_j$ for every $j \in J$.

When w is not stored in $\overrightarrow{\mathbf{w}\omega}$, the situation is a bit more complicated. Let $\widetilde{\sigma}_0(w) = w$, $\widetilde{\sigma}_{k+1}(w) = \sigma_k(w)\sigma(\sigma_k(w))$, be the ever longer sequence of finite prefixes of $\widetilde{\sigma}(w)$. If no $\widetilde{\sigma}_k(w)$ is stored in $\overrightarrow{\mathbf{w}\omega}$, then $\widetilde{f_{\overrightarrow{\mathbf{w}\omega}}}(w) = \widetilde{\sigma}(w)$. But there may be a $k \in \mathbb{N}$ such that $\widetilde{\sigma}_k(w)$ is stored in $\overrightarrow{\mathbf{w}\omega}$. Taking the least such k, it must be the case that $\widetilde{f_{\overrightarrow{\mathbf{w}\omega}}}(w) = w_j\omega_j$ where j is any index of J such that $w_j \preceq \widetilde{\sigma}_k(w) \preceq w_j\omega_j$. These remarks being made, let us proceed.

We wish to show that the map $\alpha_{\mathcal{W}} \colon \overrightarrow{\mathbf{w}\omega} \in Cst_{\mathcal{W}} \mapsto f_{\overrightarrow{\mathbf{w}\omega}} \in \widetilde{Sel}(F)$ is measurable. For now, fix $w \in Z\Lambda^*$.

The set S_k of tuples $\overrightarrow{\mathbf{w}\omega} \in Cst$ such that $\widetilde{\sigma}_k(w)$ is stored in $\overrightarrow{\mathbf{w}\omega}$, is measurable. Indeed, $\widetilde{\sigma}_k(w)$ is stored in $\overrightarrow{\mathbf{w}\omega}$ if and only if there is a $j \in J$ such that $w_j \preceq \widetilde{\sigma}_k(w) \preceq w_j\omega_j$, so $S_k = \bigcup_{\substack{j \in J \\ w_j \preceq \widetilde{\sigma}_k(w)}} \pi_j^{-1}(\widetilde{\sigma}_k(w)\Lambda^{\mathbb{N}})$, where $\pi_j \colon Cst \to Path_F(w_i)$ is projection onto the jth component. It follows that the set $S_{=k} = S_k \smallsetminus \bigcup_{\ell=0}^{k-1} S_\ell$ of tuples $\overrightarrow{\mathbf{w}\omega} \in Cst$ such that k is the least natural number such that $\widetilde{\sigma}_k(w)$ is stored in $\overrightarrow{\mathbf{w}\omega}$ is also measurable. Let S_∞ be the (measurable) complement of $\bigcup_{k=0}^{+\infty} S_k$.

Let φ_k map each $\overrightarrow{\mathbf{w}\omega} \in S_k$ to $\widetilde{f_{\overrightarrow{\mathbf{w}\omega}}}(w)$, i.e., to $w_j\omega_j$ where $j \in J$ is such that $w_j \preceq \widetilde{\sigma}_k(w) \preceq w_j\omega_j$. For every measurable subset $\mathcal{E}$ of $Z\Lambda^{\mathbb{N}}$, $\varphi_k^{-1}(\mathcal{E}) = \bigcup_{\substack{j \in J \\ w_j \preceq \widetilde{\sigma}_k(w)}} \widetilde{\sigma}_k(w)\Lambda^{\mathbb{N}}$ is measurable, so φ_k is measurable. Also, the function φ_∞ that maps each $\overrightarrow{\mathbf{w}\omega} \in S_\infty$ to $\widetilde{f_{\overrightarrow{\mathbf{w}\omega}}}(w) = \widetilde{\sigma}(w)$ is measurable since constant. By patching φ_k, $k \in \mathbb{N}$, and φ_∞ (Lemma 4), we obtain that the map $\alpha_w \colon \overrightarrow{\mathbf{w}\omega} \in Cst \mapsto \widetilde{f_{\overrightarrow{\mathbf{w}\omega}}}(w)$ is measurable.

We now observe that, for all $w \in Z\Lambda^*$ and $\mathcal{E} \in \Sigma_{\Lambda^{\mathbb{N}}}$, $\alpha_{\mathcal{W}}^{-1}([w \xrightarrow{\sim} w\mathcal{E}]) = \{\overrightarrow{\mathbf{w}\omega} \in Cst_{\mathcal{W}} \mid \widetilde{f_{\overrightarrow{\mathbf{w}\omega}}}(w) \in w\mathcal{E}\} = \alpha_w^{-1}(w\mathcal{E})$, which is measurable: so $\alpha_{\mathcal{W}}$ is measurable. □