Responsibility Attribution in Parameterized Markovian Models

Christel Baier[1], Florian Funke[1], and Rupak Majumdar[2]
[1]Technische Universität Dresden, Dresden, Germany
[2]MPI-SWS, Kaiserslautern, Germany
{christel.baier, florian.funke}@tu-dresden.de, rupak@mpi-sws.org

Abstract

We consider the problem of responsibility attribution in the setting of parametric Markov chains. Given a family of Markov chains over a set of parameters, and a property, responsibility attribution asks how the difference in the value of the property should be attributed to the parameters when they change from one point in the parameter space to another.

We formalize responsibility as *path-based* attribution schemes studied in cooperative game theory. An attribution scheme in a game determines how a value (a surplus or a cost) is distributed among a set of participants. Path-based attribution schemes include the well-studied Aumann-Shapley and the Shapley-Shubik schemes. In our context, an attribution scheme measures the responsibility of each parameter on the value function of the parametric Markov chain.

We study the decision problem for path-based attribution schemes. Our main technical result is an algorithm for deciding if a path-based attribution scheme for a rational (ratios of polynomials) cost function is over a rational threshold. In particular, it is decidable if the Aumann-Shapley value for a player is at least a given rational number. As a consequence, we show that responsibility attribution is decidable for parametric Markov chains and for a general class of properties that include expectation and variance of discounted sum and long-run average rewards, as well as specifications in temporal logic.

Introduction

Methods that explain the behavior of complex mathematical models has become an important research direction in recent years, as such models are increasingly used in making decisions that affect our lives in crucial ways. An important problem in explainability is *responsibility attribution*: a quantitative estimation of the relative importance of model features to a final outcome. While explainability and responsibility has been broadly studied for many statistical models, to the best of our knowledge, they have never been formalized for Markov chains.

In this paper, we develop a theory of responsiblity for parametric Markov chains (pMCs) based on attribution schemes from game theory. Our perspective on responsibility focuses on the influence of a parameter on properties in pMCs and does not relate to moral, epistemic, or organizational considerations. In an abstract setting, attribution schemes consider a function $f(x_1, \ldots, x_n)$ in n parameters. When the parameters change from one setting to another, thereby changing the value of the function, the *attribution problem* is to determine what portion of the overall change in f should be attributed to each parameter. Attribution problems formalize the informal notion of responsibility in a quantitative sense: the attribution of a specific parameter can be interpreted as its responsibility in the overall change.

Attribution problems formalize the notion of responsibility in many areas of economics and engineering. For an economics example, suppose that multiple suppliers cooperate to produce a product. The profit $f(x_1, \ldots, x_n)$ has to be shared between them. For an example pertaining to pMCs, consider a fault tree modeling the overall failure probability of an entire system based on the failure probabilities of its components. A natural question is how "responsible" each component is to the overall failure of the system; that is, how should the responsibility of failure be allocated to the failure of each individual component?

One can write down some properties that an attribution function must satisfy (the "axiomatic approach"). For example, the sum of attributions over all parameters should be the change in the value of the function, if the function is independent of a parameter then its attribution should be zero, and the attribution should not depend on the identity of the parameters. When f is a linear function, $f(x_1, \ldots, x_n) = \sum_i c_i x_i$, and each parameter changes from an initial value a_i to a final value b_i, a simple attribution scheme could assign the attribution $c_i(b_i - a_i)$ to the ith parameter. However, when f is non-linear, a small change in one parameter may be responsible for a large change in f and a linear attribution would not take this into account. When each parameter changes *infinitesimally*, one can compute the partial derivative of the function f with respect to each parameter at the current valuation; this corresponds to approximating f locally by a linear function. If the parameters change by a significant amount, the partial derivative is a poor choice.

Instead, the cooperative game theory literature, and more recently, the machine learning literature, has considered *path-based* attribution schemes (Friedman and Moulin 1999). Given an initial and a final value of the parameters, a path-based attribution scheme fixes a family of paths and assigns an attribution to a parameter by integrating the partial derivative of the function f along the paths. Path-based attribution is a very general technique; it obtains as special cases the classical values of cooperative game theory such as the Shapley-Shubik value (Shapley and Shubik 1954; Shubik 1962) and the Aumann-Shapley value (Aumann and Shapley 1974). Moreover, they can be characterized axiomatically.

We consider the decision problem for path-based attribution schemes. Given a function f, two evaluation points that give values to each parameter, and a vector of rationals r, one for each parameter, the attribution decision problem asks if a specific path-based attribution scheme assigns at least r_i to parameter i to explain the change in f between the two evaluation points. We show this problem is decidable when f is a rational function (a ratio of polynomials).

In particular, the above decision problem is decidable for Shapley-Shubik and the Aumann-Shapley values for rational functions. Decidability of the Shapley-Shubik value is not surprising: by definition, it corresponds to the sum of exponentially many evaluations of a rational function.[1] That the Aumann-Shapley value is also decidable is more subtle, since the Aumann-Shapley value involves computing definite integrals of rational functions and therefore contains transcendental functions. While one can numerically approximate such functions to arbitrary precision, one may still not be able to distinguish where the value of the value of the integral lies with respect to the rational threshold.

Our decidability result uses Baker's theorem from transcendental number theory (Baker 1977). The definite integral of a rational function along an affine path can be written as a linear form in logarithms of algebraic numbers. We use Baker's theorem, and algorithms on algebraic numbers, to provide an algorithm to check if the linear form is greater than the given rational.

Our result proves, as a special case, decidability of path-based attribution schemes for pMCs. A pMC represents a family of Markov chains, one for each choice of the parameters. We consider a broad class of specifications on pMCs, which includes classical discounted and long run average rewards, as well as specifications given by formulas of (quantitative) temporal logics. For this broad class of specifications, the expected value (and in fact the variance) can be obtained as a *rational* function of the parameters under semi-algebraic constraints. Using our decidability result, we conclude that the attribution problem for pMCs against this class of specifications is decidable.

We summarize our contributions as follows.

- We formalize the notion of responsibility for pMCs using attribution schemes from game theory;
- We provide an algorithm to decide if a path-based attribution scheme for a rational function is over a threshold;
- As a special case, we apply the algorithm to show decidability of the responsibility attribution problem for pMCs.

[1] Note that computing the Shapley value in simple settings is already #P-hard (Deng and Papadimitriou 1994).

Related Work

As far as we know, the responsibility problem has not been studied for operational stochastic models, nor was the relationship between attribution schemes and responsibility explored in this context.

The Shapley value (Shapley 1953) is a fundamental building block in the understanding of cooperative games. Generalizations of this central notion include the Shapley-Shubik value (Shapley and Shubik 1954; Shubik 1962) and the Aumann-Shapley value (Aumann and Shapley 1974). This theory was applied in the economics literature under the name of cost-sharing schemes (Mirman and Tauman 1982; Billera, Heath, and Raanan 1978; Billera and Heath 1982; Friedman and Moulin 1999), where suitable axiomatizations distill the aforementioned values as canonical schemes. Computational complexity questions for variants of the Shapley value have been studied extensively (Deng and Papadimitriou 1994; Fatima, Wooldridge, and Jennings 2008; Skibski et al. 2019, 2020) but not the Aumann-Shapley value. Recently, Shapley-like values have been rediscovered for the explanation of machine learning models (Lundberg and Lee 2017; Lundberg, Erion, and Lee 2018; Sundararajan and Najmi 2019). In this context, attribution schemes help to measure the influence of the input parameters on the outcome of the learned model.

Parametric Markov chains (pMC) have initially been introduced in a restricted form, where transition probabilities belong to certain intervals (Jonsson and Larsen 1991; Givan, Leach, and Dean 2000; Kozine and Utkin 2002). Model checking this class of pMCs against probabilistic computation tree logic (PCTL) has been considered (Sen, Viswanathan, and Agha 2006). It has also been studied how valuations for pMCs with prescribed properties can be found (Lanotte, Maggiolo-Schettini, and Troina 2007). Perturbation analysis on pMCs in the spirit of (Chen et al. 2014; Su et al. 2016) discusses how volatile a given property is under changes of the parameters. The distance-based perspective on the parameter space employed there results in a global viewpoint on parameter changes. Our approach, on the other hand, takes an individual look at the parameters and proposes a measure for their individual influence.

Techniques for computing the functions associated to PCTL specifications in pMCs has been subject to an extensive amount of research. A first approach relied on state-elimination (Daws 2005). Significant computational improvements were subsequently made (Hahn, Hermanns, and Zhang 2011) and led to the implementation PARAM (Hahn et al. 2010), as well as a reimplementation in the model checker PRISM (Kwiatkowska, Norman, and Parker 2011). Further technical insights on the efficient computation of these functions (Jansen et al. 2014) resulted in the model checker STORM (Dehnert et al. 2017). Recently, fraction-free Gaussian elimination was employed to speed up the calculation of value functions (Baier et al. 2020). Laplace expansion has been applied to solve the linear equation systems of PCTL specifications in sparse pMCs (Filieri, Ghezzi, and Tamburrelli 2011).

Finally, responsibility as a quantitative measure of blame has a rich history in the causality literature (Chockler and Halpern 2004; Aleksandrowicz et al. 2014; Chockler 2016; Alechina, Halpern, and Logan 2020; Friedenberg and Halpern 2019). Although such models have been studied for non-probabilistic Kripke structures (see, e.g., (Bulling and Dastani 2013; Chockler 2016; Yazdanpanah and Dastani 2016; Yazdanpanah et al. 2019)), they had not been applied to pMCs nor formulated as path-based attribution problems. In philosophy (van de Poel 2011), one distinguishes between forward responsibility, which is a global notion of responsibility for the entire model, and backward responsibility, which is assigned relative to a specific unfolding of events. Our approach falls into the second category since we attribute responsibility values after observing a change in the parameter setting.

Preliminaries

We write $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ for the naturals, rationals, reals, and complex numbers, respectively. For a field K and variables $X = \{x_1, ..., x_n\}$, we write $K[X]$ for the polynomial ring over X and $K(X)$ for the field of rational functions over X. Given elements $f \in K[X]$ and $c \in \mathbb{R}^X$ we denote by $f[c] \in \mathbb{R}$ the value obtained by evaluating f at c. A *partial function* $f\colon \mathbb{R}^X \dashrightarrow \mathbb{R}$ is a function $D \to \mathbb{R}$ on some subset $D \subseteq \mathbb{R}^X$. A rational function $f = g/h \in K(X)$ naturally induces a partial function (written by the same letter) $f\colon \mathbb{R}^X \dashrightarrow \mathbb{R}$ defined on $D = \{c \in \mathbb{R}^X \mid h[c] \neq 0\}$ by evaluating numerator and denominator.

A function $f(x_1, ..., x_n)\colon D \to \mathbb{R}$ is *independent* of the ith variable if $f(x) = f(x')$ whenever $x_j = x'_j$ for all $j \neq i$. It is *non-decreasing* in the ith variable if $f(x + t\delta_i) \geq f(x)$ is non-decreasing in t, where $\delta_i = (0, ..., 0, 1, 0, ..., 0)$ is the vector containing the 1 in the ith position. An *entrywise affine* map is a map of the form

$$h\colon \mathbb{R}^X \to \mathbb{R}^X, (x_1, ..., x_n) \mapsto \left(\frac{x_1 - b_1}{a_1}, ..., \frac{x_n - b_n}{a_n}\right)$$

for some $a_i, b_i \in \mathbb{R}, a_i \neq 0$.

Definition 1 (Parametric Markov chain)**.** *A* parametric Markov chain (pMC) $M = (S, \mathrm{AP}, L, s_0, X, P, R)$ *consists of a finite set of states S, atomic propositions* AP*, a labeling $L\colon S \to 2^{\mathrm{AP}}$, an initial state $s_0 \in S$, a finite set X of parameters, a parametric probabilistic transition function $P : S \times S \to \mathbb{Q}(X)$, a parametric state reward function $R\colon S \to \mathbb{Q}(X)$.*

A valuation $c \in \mathbb{R}^X$ is *admissible* for a pMC M if evaluating all parametric inputs at c results in a Markov chain with state rewards, i.e., we have $0 \leq P(s,t)[c] \leq 1$ for all $s, t \in S$, $\sum_{t\in S} P(s,t)[c] = 1$ for all $s \in S$ and $R(s)[c]$ is defined. A set $D \subseteq \mathbb{R}^X$ is *admissible* if all valuations in D are admissible.

Reasoning about responsibility typically involves *agents* that have control over certain *actions*. Agency in the pMC models comes from the range of parameters: for each x_i, we assume an independent agent can perform a change in value of the parameter x_i independently from other parameters.

Definition 2 (Regular property)**.** *A* property *(over* AP*) is a map that assigns to each pMC $M = (S, \mathrm{AP}, L, s_0, X, P, R)$ a partial function $\phi_M : \mathbb{R}^n \dashrightarrow \mathbb{R}$ defined on the set of admissible valuations for M. A property is* regular *if for all pMCs M the function ϕ_M is induced by a rational function in $\mathbb{Q}(X)$*

Our notion of a regular property is designed in such a general way that virtually all properties of major interest in classical Markov chain theory as well as in probabilistic model checking fit into this framework. We now recall our main examples. Let $M = (S, \mathrm{AP}, L, s_0, P, R)$ be a Markov chain, i.e., a pMC without parameters, where P and R map to $\mathbb{Q}$ instead of $\mathbb{Q}(X)$. An infinite path in M is a sequence $s_0 s_1 s_2 \ldots$ such that $P(s_i, s_{i+1}) > 0$ for all $i \geq 0$. The set of infinite paths $\mathrm{Paths}(M)$ can be turned into a probability space via a standard cylinder construction (cf. (Baier and Katoen 2008, Section 10.1)). Given an ω-regular property over AP, we write $\Pr_M(\phi)$ short for the mass of those paths in M starting in s_0 satisfying ϕ. We also incorporate Probabilistic Computation Tree Logic (PCTL). This branching-time logic is formed according to the grammar

$$\begin{aligned} \Phi \ &::= \ \mathrm{true} \ \mid \ a \ \mid \ \Phi_1 \wedge \Phi_2 \ \mid \ \neg\Phi \ \mid \ \mathbb{P}_J(\phi) \\ \phi \ &::= \ \bigcirc\Phi \ \mid \ \Phi_1 \,\mathsf{U}\, \Phi_2 \end{aligned}$$

where $a \in \mathrm{AP}$, and J is an interval in $[0, 1]$. A formula of the form Φ is called *state formula* and a formula of the form ϕ is called *path formula*. The operators $\bigcirc$ (next) and U (until) denote the usual temporal modalities. A state s satisfies the state formula $\mathbb{P}_J(\phi)$ if the mass of those paths in M starting in s satisfying ϕ belongs to the interval J.

Consider a target set $T \subseteq S$ with $\Pr_M(\Diamond T) = 1$. Then the state reward function R induces together with a discount factor $0 < t \leq 1$ a random variable on infinite paths defined by mapping the path $s_0 s_1 ...$ to the value $\sum_{i=0}^{n} t^i \cdot R(s_i)$, where n is the smallest index such that $s_n \in T$. The set of paths for which such an n does not exists is a null set since $\Pr_M(\Diamond T) = 1$. The *expected discounted reward* $\mathbb{ER}_{M,t}(\Diamond T)$ is the expected value of this random variable, and the *variance of rewards* $\mathrm{Var}\,\mathbb{R}_{M,t}(\Diamond T)$ is its variance.

Lemma 1. *The following are regular properties on pMCs $M = (S, \mathrm{AP}, L, s_0, X, P, R)$:*

1. ω-regular properties: $\phi_M(c) = \Pr_{M[c]}(\phi)$*, where ϕ is an ω-regular property over* AP*;*
2. PCTL path properties: $\phi_M(c) = \Pr_{M[c]}(\phi)$*, where ϕ is a PCTL path property over* AP*;*
3. Expected discounted reward: $\phi_M[c] = \mathbb{ER}_{M[c],t}(\Diamond T)$*, where $T \subseteq S$ and $0 < t \leq 1$;*
4. Variance of rewards*:* $\phi_M[c] = \mathrm{Var}\,\mathbb{R}_{M,t}(\Diamond T)$*, where $T \subseteq S$ and $0 < t \leq 1$.*

Proof. The arguments rely on well-known non-parametric constructions. For an ω-regular property, we follow (Baier and Katoen 2008, Section 10.3): One takes a deterministic Rabin automaton $\mathcal{A}$ for ϕ and builds the product of $M \otimes \mathcal{A}$. This is a pMC with the same parameter set as M and satisfaction of ϕ in M translates into reachability of a subset of the bottom strongly connected components of $M \otimes \mathcal{A}$. These resulting parametric reachability probabilities are the unique solution of a linear equation system $Ax = b$ over $\mathbb{Q}(X)$, where A and b contain essentially the probability function P, cf. (Baier and Katoen 2008, Theorem 10.19). As the inverse of a matrix over $\mathbb{Q}(X)$ depends rationally upon its entries, the solution $x = A^{-1}b$ depends rationally on P and thus on X.

For PCTL we consider the two cases $\phi = \bigcirc\Phi$ and $\phi = \Phi_1 \cup \Phi_2$ separately. In the first case, one obtains the vector of probabilities $(\Pr_{M[c],s}(\phi))_{s\in S}$ by multiplying the parametric probability transition function P with the vector over $\mathbb{R}^S$ containing a 1 for all states satisfying Φ and a 0 for the remaining states. In the second case where $\phi = \Phi_1 \cup \Phi_2$ one builds a fixed point linear equation system in the very same fashion as for the reachability probabilities considered above. Further details can, for example, be found in (Baier and Katoen 2008, Section 10.2.1).

Similarly, one can describe a linear equation system $A'x = b'$ over $\mathbb{Q}(X)$ whose only solution is the expected discounted reward $\mathbb{ER}_{M[c],t}(\Diamond T)$, cf. (Baier and Katoen 2008, Section 10.5). An analogous approach has been identified for variances of rewards $\operatorname{Var}\mathbb{R}_{M,t}(\Diamond T)$ (Verhoeff 2004). In both cases the non-parametric approach naturally extends to pMCs by solving the linear equation systems over $\mathbb{Q}(X)$ instead of $\mathbb{Q}$. □

Nevertheless, an explicit computation of the rational function that belongs to a regular property is a non-trivial task since solving linear equation systems over rational function fields is much harder than over $\mathbb{Q}$. Moreover, even for simple examples the resulting function might require a monomial representation of exponential size. More precisely, there are sequences $(M_k)_{k\geq 1}$ of acyclic pMCs with k parameters and $k+3$ states (one of which is a goal state g) such that P can be described by linear functions and the reachability probability $\Pr_{M[\cdot]}(\Diamond g)$ is a polynomial with 2^k monomials (Baier et al. 2020).

Example 1 (Parameterized Knuth's Dice). Knuth's dice is a Markov chain, due to Knuth and Yao (Knuth and Yao 1976), which models a fair dice using only fair coins. We consider a version of the program where we simulate a biased dice using three biased coins, whose probabilities of heads are parameterized by x, y, and z (see Figure 1). Starting at the initial state s_0 of the pMC, we can write down rational functions for the probability that the dice rolls a specific number. In PCTL notation, the probability that the dice rolls a specific number is written $\Pr_{s_0}(\Diamond 1)$, $\Pr_{s_0}(\Diamond 2)$, etc. These probabilities can be calculated as follows. Note that $\Pr_{s'_{1,2,3}}(\Diamond 1) = (1-z) + zy \cdot \Pr_{s'_{1,2,3}}(\Diamond 1)$, and so $\Pr_{s'_{1,2,3}}(\Diamond 1) = \frac{1-z}{1-zy}$. Similarly, $\Pr_{s'_{1,2,3}}(\Diamond 2) = z^2(1-y) + zy \cdot \Pr_{s'_{1,2,3}}(\Diamond 2)$, and so $\Pr_{s'_{1,2,3}}(\Diamond 2) = \frac{z^2(1-y)}{1-zy}$. Thus,

$$\Pr_{s_0}(\Diamond 1) = \frac{xy(1-z)}{1-zy}$$

$$\begin{aligned}\Pr_{s_0}(\Diamond 2) &= \frac{xyz^2(1-y)}{1-zy} + x(1-y)z \\ &= \frac{xyz^2(1-y) + (1-zy)x(1-y)z}{1-zy} = \frac{x(1-y)z}{1-zy}\end{aligned}$$

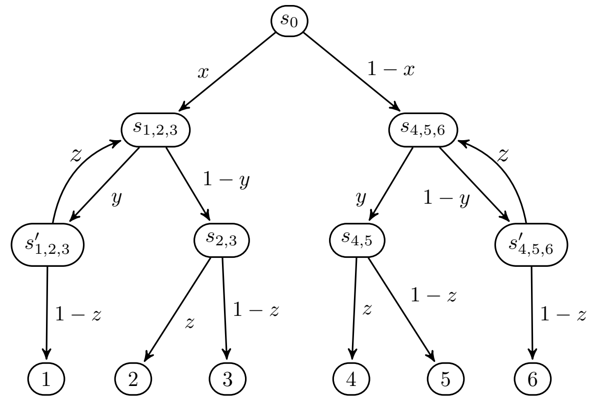

Figure 1: Knuth's dice, manipulated

Attribution Schemes

Consider a pMC $M = (S, \mathrm{AP}, L, s_0, X, P, R)$, a property ϕ, and two admissible valuations x, x' for M. In practice, when the pMC M models an engineering artifact, a designer or user has expectations on the behavior of ϕ. When dealing with network protocols or scheduler optimization tasks, for example, one typically imposes upper bounds on the fault rate or the average time until completion of the task. Imagine that, for two settings x and x' of the parameters, we see that $\phi_M(x)$ has the desired behavior but $\phi_M(x')$ does not. The *responsibility attribution problem* asks, how should the change in each parameter from x to x' be held *responsible* for the change in ϕ_M?

We tackle this problem with insights from the cost-sharing literature in economics (Mirman and Tauman 1982; Billera, Heath, and Raanan 1978; Billera and Heath 1982; Friedman and Moulin 1999), where the cost of jointly producing a good needs to be distributed among the participating suppliers. The resulting *cost-sharing schemes* have been generalized to more abstract settings (Sun and Sundararajan 2011) in the form of *attribution schemes*. We adapt this notion and apply it to regular properties on pMCs. Intuitively, an attribution scheme takes the property ϕ and divides the overall change $\phi_M(x') - \phi_M(x)$ to each of the parameters. Thus, the attribution measures the "responsibility" in producing the value $\phi_M(x')$ from the value $\phi_M(x)$.

We emphasize that our notion of responsibility is free of any moral connotation and epistemic considerations in that we focus on measuring the influence of the parameters in changing potential outcomes. Moreover, our model assumes perfect information, which is reflected by the fact that one has explicit representations for the parametric transition probabilities rather than just a representation of the rational function associated to a regular property.

Definition 3 (Attribution scheme). *Let $C^1(D)$ be the set of continuously differentiable functions $D \to \mathbb{R}$. An* attribution scheme *is a map*

$$v\colon C^1(D)\times D\times D \to \mathbb{R}^n,\ (f,x,x') \mapsto (v_i(f,x,x'))_{1\leq i\leq n}$$

such that the following properties hold:

Efficiency: *For all $x, x' \in D$ we have*

$$\sum_{i=1}^{n} v_i(f, x, x') = f(x') - f(x);$$

Dummy: *If f is independent of i, then $v_i(f, x, x') = 0$ for all $x, x' \in D$;*

Symmetry: *If $\sigma\colon \{1, ..., n\} \to \{1, ..., n\}$ is a permutation of the indices and we define*

$$f_\sigma(x_1, ..., x_n) = f(x_{\sigma(1)}, ..., x_{\sigma(n)}),$$

and likewise $f_{\sigma^{-1}}$, then we have for all $x, x' \in D$

$$v_{\sigma^{-1}(i)}(f_{\sigma^{-1}}, f_\sigma(x), f_\sigma(x')) = v_i(f, x, x')$$

Linearity: *If $f = a_1 \cdot f_1 + a_2 \cdot f_2$ for $a_i \in \mathbb{R}$ and $f_i \in C^1(D)$, then for all $x, x' \in D$*

$$v_i(f, x, x') = a_1 \cdot v_i(f_1, x, x') + a_2 \cdot v_i(f_2, x, x')$$

Non-negativity: *If f is non-decreasing in i and $x'_i \geq x_i$, then $v_i(f, x, x') \geq 0$.*

Affine Scale Invariance: *Let $h\colon \mathbb{R}^X \to \mathbb{R}^X$ be an entry-wise affine map, then for all $x, x' \in h^{-1}(D)$*

$$v_i(f \circ h, x, x') = v_i(f, h(x), h(x'))$$

The first three axioms present the basic features of the attribution problem. Efficiency states that the entire change from $f(x)$ to $f(x')$ is accounted for. Dummy demands that parameters without effect on f should not be held responsible. Symmetry is a fairness condition among the parameters: it states that the identity of a parameter should not matter in an attribution. The final three axioms impose consistency axioms on a higher level and are sometimes (partially) removed in axiomatic approaches. Linearity demands that an underlying additive structure on the functions be preserved in the attributions, while non-negativity states that parameters that cannot decrease the function in the direction from x to x' do not receive negative attributions. Affine scale invariance requires that rescaling the parameters individually does not affect the attributions.

Example 2. 1. The *Aumann-Shapley attribution* (Aumann and Shapley 1974) is a popular attribution method in the cost-sharing literature. For $x, x' \in D$ and $1 \leq i \leq n$ it is defined as

$$\mathrm{AS}_i(f, x, x') = (x'_i - x_i) \cdot \int_0^1 \frac{\partial f}{\partial x_i}(x + \alpha(x' - x))d\alpha$$

Intuitively, Aumann-Shapley breaks up the infinitesimal change on f into the individual coordinates and then accumulates these values parameter-wise along the straight line from x to x'. It is not hard to check that the axioms of Definition 3 hold.

2. The *Shapley-Shubik attribution* (Shapley and Shubik 1954) builds on the classical Shapley value from cooperative games. It is defined for two parameters $x, x' \in D$ as follows. For a set $J \subseteq \{1, ..., n\}$ we denote by $x_J x'$ the vector in D which coincides with x' on indices in J and with x on indices in the complement of J. Thus the points of the form $x_J x'$ are precisely the vertices of the hyperrectangle spanned by x and x'. We put $c(J) = f(x_J x')$, so in particular $c(\emptyset) = f(x)$ and $c(\{1, ..., n\}) = f(x')$. Then the *Shapley-Shubik attribution* $\mathrm{ShS}_i(f, x, x')$ is defined as the value

$$\sum_{J \subseteq I \setminus \{i\}} \frac{|J|!(n - |J| - 1)!}{n!} \cdot (c(J \cup \{i\}) - c(J)),$$

i.e., by forming the classical Shapley value on the payoff function c.

We have adapted the set of axioms of Definition 3 in order to fit into our framework of regular properties on pMCs. The first three axioms and non-negativity are rather natural and prevent inconsistent or counterintuitive attributions.

As for linearity, imagine that we are interested in the probability $\mathrm{Pr}_M(\Diamond(C_1 \cup C_2))$ to reach the disjoint union of two bottom strongly connected components of the pMC. This probability is the sum of the individual reachability probabilities for the two components, i.e., $\mathrm{Pr}_M(\Diamond(C_1 \cup C_2)) = \mathrm{Pr}_M(\Diamond C_1) + \mathrm{Pr}_M(\Diamond C_2)$. This inherent additive structure in the target function should be respected by our (additive) attribution scheme, meaning that the attributions for the individual attributions for the $\mathrm{Pr}_M(\Diamond C_i)$ should add up to their overall attribution. A similar reasoning also works for more complicated temporal logic formulae which can be decomposed into disjoint sets of paths, and for which the probability of satisfaction is additive.

Affine scale invariance is particularly convincing in our context. In formal modeling and in sharp contrast to the cost-sharing literature in economics, the scale of parameters does not represent real-world scales like the number of produced goods. However, if the scale of parameters in pMCs is arbitrary from the start, rescaling them individually should not affect a sensible attribution.

The cost-sharing literature also studies other axioms with which one can (at least on the class of non-decreasing functions, and for parameters $x, x' \geq 0$) uniquely characterize the attribution schems of Example 2. For the Aumann-Shapley attribution one has to add a proportionality axiom to the list of Definition 3, while for the Shapley-Shubik attribution one has to additionally require a monotonicity axiom (Friedman and Moulin 1999). We omit these axioms as they do not capture natural requirements for pMCs.

Remark 1 (Group responsibilities)**.** *It is an interesting question how attribution schemes can be used for the definition of group responsibilities, i.e., the overall responsibility of a set of parameters $\{x_i\}_{i \in I}$ for some $I \subseteq \{1, ..., n\}$. To the best of our knowledge, the corresponding problem on cooperative games and the classical Shapley value does not have an immediate answer. The* Owen value *(Owen 1977) provides an approach for* individual *responsibilities when groups have already been formed. In our context, one could define the responsibility of a set of parameters* a posteriori *as the sum of the individual responsibilities. For the Aumann-Shapley value one can see formally that this makes sense by replacing each of the parameters $x_i, i \in I$, with the multiple $(1 - z) \cdot x_i + z \cdot x'_i$ of a new parameter z. Then the Aumann-Shapley attribution of the new parameter z is the sum of the attributions of all replaced parameters.*

Path Attribution Schemes

Both the Aumann-Shapley and the Shapley-Shubik attribution fall into the general class of path attribution schemes.

Definition 4 (Path attribution scheme). *Let* $\gamma = (\gamma_1, ..., \gamma_n)\colon [0,1] \to [0,1]^n$ *be a path from* $(0, ..., 0)$ *to* $(1, ..., 1)$ *which is non-decreasing in every component. For every* $x, x' \in D$*, let*

$$\gamma_{x,x'}(t) = x + (\gamma_1(t) \cdot (x'_1 - x_1), ..., \gamma_n(t) \cdot (x'_n - x_n))$$

Denote the components of $\gamma_{x,x'}$ *by* $\gamma_{x,x',i}$. *The attribution scheme* induced by γ *is defined as*

$$v_i^\gamma(f, x, x') = \int_0^1 \frac{\partial f}{\partial x_i}(\gamma_{x,x'}(t)) \cdot \gamma'_{x,x',i}(t)\, dt$$

A path attribution scheme *is a convex combination of finitely many attributions schemes induced by paths.*

What we term *path attribution* is sometimes called *affine path attribution* in order to highlight that the paths $\gamma_{x,x'}$ are affine images of a common path on the unit cube. One could abandon this uniformity condition at the cost of giving up affine scale invariance. Our decidability results in the next section apply to that setting as well, but we maintain affine scale invariance as that is a reasonable axiom for responsibility attribution of pMCs.

Example 3. The Aumann-Shapley attribution is induced from the straight path $\gamma\colon [0,1] \to [0,1]^n, t \mapsto (t, ..., t)$. The Shapley-Shubik attribution is the (uniform) convex combination of the $n!$ many paths on the unit cube $[0,1]^n$ that successively switch the entries from 0 to 1. This can most easily be seen from the permutation-based formulation of the Shapley value (Shapley 1953). More generally, any convex combination of these $n!$ many paths is called a *probabilistic value* (Weber 1988) or *random order value* (Friedman and Moulin 1999).

The following lemma follows easily from the basic properties of integration and differentiation, see also (Sun and Sundararajan 2011).

Lemma 2. *Any affine path attribution scheme is an attribution scheme, i.e., it satisfies the six axioms of Definition 3.*

Example 4. For parameterized Knuth's dice, we compute the Aumann-Shapley value as follows. The partial derivatives are

$$\frac{\partial \mathrm{Pr}_{s_0}(\Diamond 1)}{\partial x} = \frac{y(1-z)}{1-zy},$$
$$\frac{\partial \mathrm{Pr}_{s_0}(\Diamond 1)}{\partial y} = \frac{x(1-z)}{(1-zy)^2},$$
$$\frac{\partial \mathrm{Pr}_{s_0}(\Diamond 1)}{\partial z} = \frac{xy(y-1)}{(1-zy)^2}.$$

Let us consider $p = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ as the baseline valuation and $q = (\frac{3}{4}, \frac{1}{4}, \frac{1}{2})$ as a second valuation. For the target function we then have $\mathrm{Pr}^p_{s_0}(\Diamond 1) = 1/6$ and $\mathrm{Pr}^q_{s_0}(\Diamond 1) = 3/28$. Computing the Aumann-Shapley attribution yields

$$\begin{aligned}
&\mathrm{AS}_x(\mathrm{Pr}_{s_0}(\Diamond 1), p, q)\\
&\quad= \left(\frac{3}{4} - \frac{1}{2}\right) \cdot \int_0^1 \frac{\frac{1}{4} + \frac{1}{4}t}{2 - (\frac{1}{4} + \frac{1}{4}t)}\, dt\\
&\quad= \frac{1}{4} \cdot \int_0^1 \frac{1+t}{7-t}\, dt = 0.0583\\
&\mathrm{AS}_y(\mathrm{Pr}_{s_0}(\Diamond 1), p, q)\\
&\quad= \left(\frac{1}{4} - \frac{1}{2}\right) \cdot \int_0^1 \frac{\frac{1}{2} \cdot (\frac{3}{4} - \frac{1}{4}t)}{(1 - \frac{1}{2}(\frac{1}{4} + \frac{1}{4}t))^2}\, dt\\
&\quad= -\int_0^1 \frac{6-2t}{(7-t)^2}\, dt = -0.1178\\
&\mathrm{AS}_z(\mathrm{Pr}_{s_0}(\Diamond 1), p, q) = 0
\end{aligned}$$

Note that the parameters x and y have opposite influence on $\mathrm{Pr}_{s_0}(\Diamond 1)$: the change in x from $1/2$ to $3/4$ has a positive influence on $\mathrm{Pr}_{s_0}(\Diamond 1)$, while the change in y from $1/2$ to $1/4$ has a negative influence. Quantitatively, the change in y carries roughly twice as much weight, resulting in an overall decrease of the probability to reach 1.

We can also compute the Shapley-Shubik value from the definition. We write $c(S) = \mathrm{Pr}_{s_0}(\Diamond 1)(p_S q)$ for the value of the target function on points of the form $p_S q$ which assign values from p to parameters in S and value from q to the other parameters. We have $c(\emptyset) = \frac{1}{6}$, $c(\{x\}) = \frac{1}{4}$, $c(\{x,y\}) = \frac{3}{28}$, $c(\{x,z\}) = \frac{1}{4}$, $c(\{y\}) = \frac{1}{14}$, $c(\{z\}) = \frac{1}{6}$, $c(\{y,z\}) = \frac{1}{14}$, and $c(\{x,y,z\}) = \frac{3}{28}$. We omit the calculations, but obtain

$$\mathrm{ShS}_x = \frac{5}{84} = 0.0595, \quad \mathrm{ShS}_y = -\frac{5}{42} = -0.1190,$$

and, of course, $\mathrm{ShS}_z = 0$. Qualitatively, one can explain the attributions as for Aumann-Shapley values, but note that the actual values are slightly different.

Lemma 2 ensures that there is a large number of attribution schemes. However, in specific use cases particular path-based schemes might be more suitable than others, For example, in some situations it is unrealistic that all parameters can be changed simultaneously (as necessary for the Aumann-Shapley attribution), and in other situations the overall change of a parameter cannot be performed in one step (as necessary for the Shapley-Shubik attribution). The paths inducing the attribution schemes should reflect 'admissible' real-world behavior. On the other hand, if no a priori domain knowledge is given, using the Aumann-Shapley and the Shapley-Shubik scheme gives theoretically sound attributions that can serve as preliminary results.

Decision Problem

We associate a natural decision problem with path attribution schemes.

Definition 5 (Path attribution decision problem). *Given a function* $f : \mathbb{R}^n \to \mathbb{R}$*, a path attribution scheme* γ*, two parameter valuations* $x, x' \in \mathbb{Q}^n$*, and a rational vector* $r \in \mathbb{Q}^n$*, the* path attribution decision problem *asks if* $v_i^\gamma(f, x, x') \geq r_i$ *for each* $i \in \{1, \ldots, n\}$.

In order to make the decision problem more precise, we focus on piecewise linear paths specified by rational endpoints. Both the Shapley-Shubik and the Aumann-Shapley schemes are piecewise linear attribution schemes. From the definition of the Shapley-Shubik scheme, it is obvious that it is decidable when f is a rational function. In fact, computing the Shapley-Shubik scheme involves evaluating a rational function in $\mathbb{Q}(X)$ at exponentially many different points. However, the Aumann-Shapley value cannot be computed exactly: the definite integral of a rational function can involve transcendental functions. Therefore, it is not obvious that the path attribution decision problem is decidable in a more general case.

Decidability of Path Attributions

In this section we show:

Theorem 1. *Given a pMC M over parameters X, a regular property φ, two admissible parameter valuations x, x' for M, an index i, a rational number $r \in \mathbb{Q}$, and a piecewise linear path attribution scheme v, it is decidable if $v_i(\varphi_M, x, x') \geq r$.*

Theorem 1 follows from the following general result.

Theorem 2. *Let $p, q \in \mathbb{Q}[x]$ be univariate polynomials such that q does not have a zero in $[0,1]$. Given $r \in \mathbb{Q}$, it is decidable if*

$$\int_0^1 \frac{p(x)}{q(x)} dx \geq r \tag{1}$$

The proof of the theorem is organized as follows. Using elementary calculus, we know that the integral can be written as a linear form in logarithms of algebraic numbers (Lemma 3). Then, using Baker's theorem from transcendental number theory, we can show that the linear form is either zero or transcendental and bounded away from zero. Third, using tools from computational algebraic number theory (Cohen 1993), we can compute arbitrary numerical approximations for the linear forms. Together, these steps allow us to check if the integral is above or below the rational number. In the following, we write $\overline{\mathbb{Q}}$ for the set of (possibly complex) algebraic numbers. The proof of the following lemma can be found in the appendix.

Lemma 3 (Integrals of rational functions)**.** *Let $p, q \in \mathbb{Q}[x]$ be univariate polynomials. Assume q does not have a zero in $[0,1]$. Then $\int_0^1 \frac{p(x)}{q(x)} dx$ can be written as a linear form $\beta_0 + \beta_1 \log \alpha_1 + \ldots \beta_k \log \alpha_k + \beta_{k+1} \arctan \alpha_{k+1} + \ldots + \beta_m \arctan \alpha_m$, where $\alpha_i, \beta_i \in \overline{\mathbb{Q}}$ are real algebraic numbers. Equivalently, it can be written as $\beta_0 + \beta_1 \log \gamma_1 + \ldots + \beta_m \log \gamma_1$, for $\beta_i, \gamma_i \in \overline{\mathbb{Q}}$.*

Proof (Sketch). The proof is basic calculus. First, using the fundamental theorem of algebra, $q(x)$ can be written as:

$$K \cdot \prod_{i=1}^{k} (x + \alpha_i)^{s_i} \cdot \prod_{i=1}^{l} ((x + \gamma_i)^2 + \beta_i^2)^{t_i}$$

for positive integers k, l, sequences of positive integers $(s_i)_{i=1}^k$, $(t_i)_{i=1}^l$, algebraic numbers $(\alpha_i)_{i=1}^k$, $(\beta_i, \gamma_i)_{i=1}^l$, such that all α_i are distinct and all pairs (β_i, γ_i) are distinct. The factors $(x+\alpha_i)^{s_i}$ and $((x+\gamma_i)^2+\beta_i^2)^{t_i}$ are mutually relatively prime. So, by the Euclidean algorithm, we can write

$$\frac{p(x)}{q(x)} = p(x) \cdot \left(\sum_{i=1}^{k} \frac{u_i(x)}{(x + \alpha_i)^{s_i}} + \sum_{i=1}^{l} \frac{v_i(x)}{((x + \gamma_i)^2 + \beta_i^2)^{t_i}} \right)$$

Now, each $\frac{p(x) u_i(x)}{(x+\alpha_i)^{s_i}}$ can be written as a sum of a polynomial and terms of the form $\frac{d_i}{(x+\alpha_i)^{s_i}}$. Each $\frac{p(x) v_i(x)}{((x+\gamma_i)^2+\beta_i^2)^{t_i}}$ can be written as a sum of a polynomial and terms of the form $\frac{e_i x + f_i}{(x+\alpha_i)^{s_i}}$. The integral of a polynomial is again a polynomial and the integral of the remaining fractional part can be effectively written as a sum of rational functions, logarithms of polynomials, and arctangents of polynomials over algebraic coefficients. Note that $\arctan(x) = \frac{\mathbf{i}}{2}(\log(1 - \mathbf{i}x) - \log(1 + \mathbf{i}x))$ is a logarithmic form. $\square$

Computations with Algebraic Numbers

In our algorithm, we have to compute with algebraic numbers. As irrational numbers such as $\sqrt{2} + \sqrt{3}$ or $1 + \sqrt{2}i$ are algebraic numbers, we cannot expect a finite representation for them. However, we can use tools from computational algebraic number theory to represent algebraic numbers and perform computations with them. We recall the basics (cf. (Cohen 1993)).

The *height* of a univariate polynomial $p \in \mathbb{Z}[x]$ with integer coefficients is the maximum magnitude of its coefficients. A complex number α is *algebraic* if it is the root of a univariate polynomial with integer coefficients. The defining polynomial of α, denoted p_α, is the unique polynomial of least degree, and whose coefficients do not have common factors, which vanishes at α. The degree and height of α are respectively those of p_α.

A standard representation for algebraic numbers encodes the number α as a tuple consisting of its defining polynomial together with rational approximations of its real and imaginary parts of sufficient precision to distinguish α from the other roots of p_α. More precisely, α is represented (not necessarily uniquely) by $(p_\alpha, a, b, r) \in \mathbb{Z}[x] \times \mathbb{Q}^3$ such that α is the unique root of p_α inside the circle in $\mathbb{C}$ of radius r centred at $a + b\mathbf{i}$. Given a polynomial $p \in \mathbb{Z}[x]$, it is well-known how to compute standard representations of each of its roots in time polynomial in $\|p\|$ (Cohen 1993). From now on, when referring to computations on algebraic numbers, we implicitly refer to their standard representations.

Baker's Theorem

We need the following quantitative version of Baker's theorem from transcendental number theory (Ram Murty and Rath 2014, Chapter 19).

Theorem 3 (Baker's Theorem)**.** *Let $\alpha_1, \ldots, \alpha_m$ be non-zero algebraic numbers with degrees at most d and heights at most A. Let $\beta_0, \ldots, \beta_m$ be non-zero algebraic numbers with degrees at most d and heights at most $B \geq 2$. Then either*

$$\Lambda := \beta_0 + \beta_1 \log \alpha_1 + \ldots \beta_m \log \alpha_m$$

equals zero or $|\Lambda| > B^{-C}$ *where* C *is an effectively computable number depending on* m*,* d*,* A*, and the values of the logarithms.*

A corollary of Baker's theorem is that, for algebraic numbers $\alpha_1, \ldots, \alpha_m$, and $\beta_1, \ldots, \beta_m$, the linear form $\beta_1 \log \alpha_1 + \ldots \beta_m \log \alpha_m$ is either zero or transcendental. In particular, if the linear form is nonzero, it is not rational. In order to check if a linear form is greater than or equal to a rational, we therefore need to compute it to sufficient bits of precision—we know that after the bound given in Baker's theorem, we can distinguish the linear form from the rational number.

Our last ingredient is to be able to compute logarithms of algebraic numbers to arbitrary bits of precision. The following theorem of Brent provides this piece. Given a real number r and a positive integer m, we say that $q \in \mathbb{Q}$ is an m-bit approximation of r if $|r - q| < 2^{-m}$.

Theorem 4. *(Brent 1976) For any fixed real numbers* $0 < a < b$*, there exists an algorithm which, given an integer* $m \geq 0$*, evaluates* $\log x$ *and* $\arctan x$ *in time* $O(m \log^2 m \log \log m)$*, with relative error* $O(2^{-m})$*, uniformly for all* $x \in [a, b]$*.*

Proofs of the Main Theorems

To check if the integral (1) is above a rational number, we compute an $\lceil -C \log B \rceil$-bit approximation of the linear form in logarithms. If the approximation is less that B^{-C}, we know that the linear form is zero. Then, we need to compare 0 to an algebraic number, which is possible through standard computations with algebraic numbers.

Suppose the linear form is not zero. From the consequence of Baker's theorem, we know that the linear form is transcendental. Then, we can compare the sign of $\beta_0 + \sum_i \beta_i \log \alpha_i - r$ by computing the number to sufficient precision.

We complete the proof of Theorem 1. As φ is a regular property, it is characterized by a rational function φ_M, together with a set of semi-algebraic constraints to maintain admissibility. In order to compute the attribution scheme, we must make sure the two valuations are admissible and the set of valuations along each piecewise linear path are all admissible. Finally, we have to make sure that the rational function is defined at all valuations along the paths. Each of the above checks are decidable, as the theory of reals is decidable.

After these checks, $v_i(\varphi_M, x, x')$ is given as a convex combination of finitely many integrals over rational functions on the inveral $[0, 1]$. Using additivity, this value can be written as one integral of a rational function on $[0, 1]$. We then use Theorem 2 to compare this integral to the given rational.

A Remark on Complexity

In the above, we only show decidability of the problem. One can get a complexity estimate by using sharper quantitative bounds in Baker's theorem (Baker 1977; Baker and Wustholz 1993). These quantitative bounds show that $\log |\Lambda|$ can be exponentially small in the parameters of the linear form, and thus, give an exponential algorithm to check if a linear form is zero or non-zero. On the other hand, we do not know any non-trivial complexity lower bounds for the problem.

Conclusion

We have provided computability results for path attribution schemes for rational cost functions. Our main technical result shows that one can decide if the Aumann-Shapley value for rational cost functions is greater than or less than a rational number. As an application of attribution schemes to pMCs, we obtain a formalization of responsibility in that domain. Moreover, since the value function is a rational function of parameters for a broad class of properties on Markov chains, we immediately get a decidability result for such properties as well. Our formalization can form the basis for responsibility analysis for application domains modeled as pMCs, such as fault trees. Finally, while the exact decision procedure seems complex, we expect that an efficient numerical evaluation of the integral will be sufficient to provide attribution values in practice.

Acknowledgements

This work was funded by DFG grant 389792660 as part of TRR 248 – CPEC (see https://perspicuous-computing.science), the Cluster of Excellence EXC 2050/1 (CeTI, project ID 390696704, as part of Germany's Excellence Strategy), DFG-projects BA-1679/11-1 and BA-1679/12-1, and the European Research Council under the Grant Agreement 610150 (http://www.impact-erc.eu/) (ERC Synergy Grant ImPACT). We thank the reviewers for their detailed and valuable comments.

References

Alechina, N.; Halpern, J. Y.; and Logan, B. 2020. Causality, Responsibility and Blame in Team Plans. *CoRR* abs/2005.10297. URL https://arxiv.org/abs/2005.10297.

Aleksandrowicz, G.; Chockler, H.; Halpern, J. Y.; and Ivrii, A. 2014. The Computational Complexity of Structure-Based Causality. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada*, 974–980. AAAI Press.

Aumann, R. J.; and Shapley, L. S. 1974. *Values of Non-Atomic Games*. Princeton University Press. URL http://www.jstor.org/stable/j.ctt13x149m.

Baier, C.; Hensel, C.; Hutschenreiter, L.; Junges, S.; Katoen, J.-P.; and Klein, J. 2020. Parametric Markov chains: PCTL complexity and fraction-free Gaussian elimination. *Information and Computation* 272: 104504. doi:https://doi.org/10.1016/j.ic.2019.104504.

Baier, C.; and Katoen, J.-P. 2008. *Principles of Model Checking (Representation and Mind Series)*. The MIT Press, Cambridge, MA.

Baker, A. 1977. The theory of linear forms in logarithms. In *Transcendence theory: Advances and Applications (Proc. Conf., Univ. Cambridge, Cambridge, 1976)*, 1–27. Academic Press.

Baker, A.; and Wustholz, G. 1993. Logarithmic forms and group varieties. *J. reine Angew. Math.* 442: 19–62.

Billera, L. J.; and Heath, D. C. 1982. Allocation of Shared Costs: A Set of Axioms Yielding a Unique Procedure. *Mathematics of Operations Research* 7(1): 32–39. URL http://www.jstor.org/stable/3689357.

Billera, L. J.; Heath, D. C.; and Raanan, J. 1978. Internal Telephone Billing Rates – A Novel Application of Non-Atomic Game Theory. *Operations Research* 26(6): 956–965. doi:https://doi.org/10.1287/opre.26.6.956.

Brent, R. 1976. Fast multiple-precision evaluation of elementary functions. *J. ACM* 23(2): 242–251. URL https://doi.org/10.1145/321941.321944.

Bulling, N.; and Dastani, M. 2013. Coalitional Responsibility in Strategic Settings. In Leite, J.; Son, T. C.; Torroni, P.; van der Torre, L.; and Woltran, S., eds., *Computational Logic in Multi-Agent Systems*, 172–189. Berlin, Heidelberg: Springer Berlin Heidelberg.

Chen, T.; Feng, Y.; Rosenblum, D. S.; and Su, G. 2014. Perturbation Analysis in Verification of Discrete-Time Markov Chains. In Baldan, P.; and Gorla, D., eds., *CONCUR 2014 – Concurrency Theory*, 218–233. Berlin, Heidelberg: Springer Berlin Heidelberg.

Chockler, H. 2016. Causality and Responsibility for Formal Verification and Beyond. In *Proceedings First Workshop on Causal Reasoning for Embedded and safety-critical Systems Technologies, CREST@ETAPS 2016, Eindhoven, The Netherlands, 8th April 2016*, volume 224 of *EPTCS*, 1–8.

Chockler, H.; and Halpern, J. Y. 2004. Responsibility and Blame: A Structural-Model Approach. *J. Artif. Intell. Res.* 22: 93–115. doi:https://doi.org/10.1613/jair.1391.

Cohen, H. 1993. *A Course in Computational Algebraic Number Theory*. Springer-Verlag.

Daws, C. 2005. Symbolic and Parametric Model Checking of Discrete-Time Markov Chains. In Liu, Z.; and Araki, K., eds., *Theoretical Aspects of Computing - ICTAC 2004*, 280–294. Berlin, Heidelberg: Springer Berlin Heidelberg.

Dehnert, C.; Junges, S.; Katoen, J.-P.; and Volk, M. 2017. A Storm is Coming: A Modern Probabilistic Model Checker. In Majumdar, R.; and Kunčak, V., eds., *Computer Aided Verification*, 592–600. Cham: Springer Intern. Publishing.

Deng, X.; and Papadimitriou, C. 1994. On the complexity of cooperative solution concepts. *Mathematics of Operations Research* 19(2): 257–266.

Fatima, S. S.; Wooldridge, M. J.; and Jennings, N. R. 2008. A linear approximation method for the Shapley value. *Artif. Intell.* 172(14): 1673–1699. doi:https://doi.org/10.1016/j.artint.2008.05.003.

Filieri, A.; Ghezzi, C.; and Tamburrelli, G. 2011. Run-Time Efficient Probabilistic Model Checking. In *Proceedings of the 33rd International Conference on Software Engineering*, ICSE '11, 341–350. New York, NY, USA: Association for Computing Machinery. doi:https://doi.org/10.1145/1985793.1985840.

Friedenberg, M.; and Halpern, J. Y. 2019. Blameworthiness in Multi-Agent Settings. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, Honolulu, Hawaii, USA*, 525–532. AAAI Press.

Friedman, E.; and Moulin, H. 1999. Three methods to share joint costs or surplus. *Journal of Economic Theory* 87(2): 275–312.

Givan, R.; Leach, S.; and Dean, T. 2000. Bounded-parameter Markov decision processes. *Artificial Intelligence* 122(1): 71 – 109. doi:https://doi.org/10.1016/S0004-3702(00)00047-3.

Hahn, E.; Hermanns, H.; and Zhang, L. 2011. Probabilistic Reachability for Parametric Markov Models. *International Journal on Software Tools for Technology Transfer* 13(1): 3–19. doi:https://doi.org/10.1007/978-3-642-02652-2_10.

Hahn, E. M.; Hermanns, H.; Wachter, B.; and Zhang, L. 2010. PARAM: A Model Checker for Parametric Markov Models. In Touili, T.; Cook, B.; and Jackson, P., eds., *Computer Aided Verification*, 660–664. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:https://doi.org/10.1007/978-3-642-14295-6_56.

Jansen, N.; Corzilius, F.; Volk, M.; Wimmer, R.; Ábrahám, E.; Katoen, J.-P.; and Becker, B. 2014. Accelerating Parametric Probabilistic Verification. In Norman, G.; and Sanders, W., eds., *Quantitative Evaluation of Systems*, 404–420. Cham: Springer International Publishing. doi:https://doi.org/10.1007/978-3-319-10696-0_31.

Jonsson, B.; and Larsen, K. G. 1991. Specification and refinement of probabilistic processes. In *Proceedings Sixth Annual IEEE Symposium on Logic in Computer Science*, 266–277.

Knuth, D.; and Yao, A. 1976. The complexity of nonuniform random number generation. In *Algorithms and Complexity: New Directions and Recent Results*. Academic Press.

Kozine, I.; and Utkin, L. 2002. Interval-Valued Finite Markov Chains. *Reliable Computing* 8: 97–113. doi:10.1023/A:1014745904458.

Kwiatkowska, M.; Norman, G.; and Parker, D. 2011. PRISM 4.0: Verification of Probabilistic Real-Time Systems. In Gopalakrishnan, G.; and Qadeer, S., eds., *Computer Aided Verification*, 585–591. Berlin, Heidelberg: Springer Berlin Heidelberg.

Lanotte, R.; Maggiolo-Schettini, A.; and Troina, A. 2007. Parametric probabilistic transition systems for system design and analysis. *Formal Asp. Comput.* 19: 93–109. doi:https://doi.org/10.1007/s00165-006-0015-2.

Lundberg, S. M.; Erion, G. G.; and Lee, S.-I. 2018. Consistent individualized feature attribution for tree ensembles. Technical Report arxiv:1802.03888.

Lundberg, S. M.; and Lee, S.-I. 2017. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NeurIPS'17, 4768–4777. Red Hook, NY, USA: Curran Associates Inc.

Mirman, L. J.; and Tauman, Y. 1982. Demand Compatible Equitable Cost Sharing Prices. *Mathematics of Operations Research* 7(1): 40–56. URL http://www.jstor.org/stable/3689358.

Owen, G. 1977. Values of Games with a Priori Unions. In Henn, R.; and Moeschlin, O., eds., *Mathematical Economics and Game Theory*, 76–88. Berlin, Heidelberg: Springer Berlin Heidelberg.

Ram Murty, M.; and Rath, P. 2014. *Transcendental numbers*. Springer.

Sen, K.; Viswanathan, M.; and Agha, G. 2006. Model-Checking Markov Chains in the Presence of Uncertainties. In Hermanns, H.; and Palsberg, J., eds., *Tools and Algorithms for the Construction and Analysis of Systems*, 394–410. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: https://doi.org/10.1007/11691372_26.

Shapley, L. S. 1953. A value for n-person games. In *Kuhn, H., Tucker, A. (Eds.), Contributions to the Theory of Games. Vol. II.*, 307–317. Princeton University Press.

Shapley, L. S.; and Shubik, M. 1954. A Method for Evaluating the Distribution of Power in a Committee System. *The American Political Science Review* 48(3): 787–792. URL http://www.jstor.org/stable/1951053.

Shubik, M. 1962. Incentives, Decentralized Control, the Assignment of Joint Costs and Internal Pricing. *Management Science* 8(3): 325–343.

Skibski, O.; Michalak, T. P.; Sakurai, Y.; Wooldridge, M. J.; and Yokoo, M. 2020. Partition decision trees: representation for efficient computation of the Shapley value extended to games with externalities. *Auton. Agents Multi Agent Syst.* 34(1): 11. doi:https://doi.org/10.1007/s10458-019-09429-7.

Skibski, O.; Rahwan, T.; Michalak, T. P.; and Wooldridge, M. J. 2019. Enumerating Connected Subgraphs and Computing the Myerson and Shapley Values in Graph-Restricted Games. *ACM Trans. Intell. Syst. Technol.* 10(2): 15:1–15:25. doi:https://doi.org/10.1145/3235026.

Su, G.; Feng, Y.; Chen, T.; and Rosenblum, D. S. 2016. Asymptotic Perturbation Bounds for Probabilistic Model Checking with Empirically Determined Probability Parameters. *IEEE Transactions on Software Engineering* 42(7): 623–639. doi:https://doi.org/10.1109/TSE.2015.2508444.

Sun, Y.; and Sundararajan, M. 2011. Axiomatic attribution for multilinear functions. *Proceedings of the 12th ACM conference on Electronic commerce - EC '11* doi: https://doi.org/10.1145/1993574.1993601.

Sundararajan, M.; and Najmi, A. 2019. The many Shapley values for model explanation. Technical Report arXiv:1908.08474.

van de Poel, I. 2011. The Relation Between Forward-Looking and Backward-Looking Responsibility. In Vincent, N. A.; van de Poel, I.; and van den Hoven, J., eds., *Moral Responsibility: Beyond Free Will and Determinism*, 37–52. Dordrecht: Springer Netherlands. doi:https://doi.org/10.1007/978-94-007-1878-4_3.

Verhoeff, T. 2004. Reward variance in Markov chains : a calculational approach. In Cleophas, L.; and Watson, B., eds., *Proceedings of the Eindhoven FASTAR Days 2004 (Eindhoven, The Netherlands, September 3-4, 2004)*, Computer Science Reports. Technische Universiteit Eindhoven.

Weber, R. 1988. *Probabilistic values for games*, 101–119. Cambridge University Press.

Yazdanpanah, V.; and Dastani, M. 2016. Distant Group Responsibility in Multi-agent Systems. In Baldoni, M.; Chopra, A. K.; Son, T. C.; Hirayama, K.; and Torroni, P., eds., *PRIMA 2016: Principles and Practice of Multi-Agent Systems*, 261–278. Cham: Springer International Publishing.

Yazdanpanah, V.; Dastani, M.; Jamroga, W.; Alechina, N.; and Logan, B. 2019. Strategic Responsibility Under Imperfect Information. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '19, 592–600. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.