Normalized Probabilistic Semantics is Not Associative

ELENA DI LAVORE, Tallinn University of Technology, Estonia
MARIO ROMÁN, Tallinn University of Technology, Estonia
MÁRK SZÉLES, Radboud University Nijmegen, The Netherlands

Normalization, $\mathbf{D}(X+1) \to \mathbf{D}(X)+1$, fails to form a distributive law but still yields a composite *magmad* — i.e., a monad without the associativity axiom. We introduce the first normalized-by-construction probabilistic semantics: the magmad of normalized distributions. Non-associativity allows this semantics to express both observations and interventions, both Pearl's and Jeffrey's updates, and both evidential and causal interpretations. We introduce a magmadic version of do-notation. We derive results from causal inference within this framework, including Pearl's front-door and back-door criteria.

1 Introduction

Normalization is essential in probabilistic inference: Bayes' rule for updating a prior with a likelihood requires rescaling to compute a valid posterior,

$$\textit{posterior} \propto \textit{likelihood} \cdot \textit{prior}.$$

However, probabilistic programming semantics is not necessarily *normalized-by-construction*: normalization typically appears as a primitive and, when updating, denotational models rely on subdistributions [5, 48] or unnormalized distributions [14, 35, 60, 71].

We introduce the first denotational normalized-by-construction probabilistic semantics: the *magmad* — or, *non-associative "monad"* — of normalized distributions. Perhaps surprisingly, non-associativity is forced and, far from being a limitation, it grants expressive power. Let us illustrate this normalized semantics and its non-associativity with two examples of increasing complexity.

1.1 Example — The Monty Hall problem

Imagine solving the famous *Monty Hall problem* [53]: in a game show, you choose from three closed doors for a chance of winning the prize behind one of them. However, after choosing, the host opens one of the empty doors and — with two closed doors remaining — invites you to switch your guess. *Should you switch?*

Let us compute.[1] We *(i)* consider a prior uniform probability that a prize is behind any of three doors (L, M, R); then, *(ii)* after choosing, say, the middle door, the host will randomly and uniformly open a door (L, M, R) that can neither be the chosen door nor the one with the prize; *(μ_1)* we multiply and *(iii)* observe that the host opens, e.g., the left door (L) precluding the other possibility (R); upon this, *(μ_2)* we renormalize and *(π)* conclude we should switch to the right door (R): it doubles our chances of getting the prize — the posterior distribution is $\frac{1}{3}[M] + \frac{2}{3}[R]$.

$$\begin{array}{ll}
(i) & \frac{1}{3}[L] + \frac{1}{3}[M] + \frac{1}{3}[R] \\
(ii) & \frac{1}{3}\left[\frac{1}{1}[LR]\right] + \frac{1}{3}\left[\frac{1}{2}[ML] + \frac{1}{2}[MR]\right] + \frac{1}{3}\left[\frac{1}{1}[RL]\right] \\
(\mu_1) & \frac{1}{3}[LR] + \frac{1}{6}[ML] + \frac{1}{6}[MR] + \frac{1}{3}[RL] \\
(iii) & \frac{1}{3}[\bot] + \frac{1}{6}\left[\frac{1}{1}[ML]\right] + \frac{1}{6}[\bot] + \frac{1}{3}\left[\frac{1}{1}[RL]\right] \\
(\mu_2) & \frac{1}{3}[ML] + \frac{2}{3}[RL] \\
(\pi) & \frac{1}{3}[M] + \frac{2}{3}[R]\,.
\end{array}$$

[1] We use *square brackets* $[-]$ to denote the elements of a formal sum. The generic distribution over a set X is written as $\lambda_1[x_1] + \cdots + \lambda_n[x_n]$, given elements $x_1, \ldots, x_n \in X$ and scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{R}^+$ adding up to the unit, $\sum_{i=1}^n \lambda_i = 1$.

Authors' Contact Information: Elena Di Lavore, Tallinn University of Technology, Estonia; Mario Román, Tallinn University of Technology, Estonia; Márk Széles, Radboud University Nijmegen, The Netherlands.

However, the *Monty Hall problem* admits another parenthesization. Let us consider everything from the action of the host a parenthesized subproblem: we now (μ_2) normalize internally before (μ_1) normalizing globally, and the posterior is $\frac{1}{2}[M] + \frac{1}{2}[R]$ — switching doors stops mattering.

$$
\begin{aligned}
&(i) && \tfrac{1}{3}[L] + \tfrac{1}{3}[M] + \tfrac{1}{3}[R] \\
&\quad (ii) && \tfrac{1}{3}\left[\tfrac{1}{1}[LR]\right] + \tfrac{1}{3}\left[\tfrac{1}{2}[ML] + \tfrac{1}{2}[MR]\right] + \tfrac{1}{3}\left[\tfrac{1}{1}[RL]\right] \\
&\quad (iii) && \tfrac{1}{3}\left[\tfrac{1}{1}[\bot]\right] + \tfrac{1}{3}\left[\tfrac{1}{2}\left[\tfrac{1}{1}[ML]\right] + \tfrac{1}{2}[\bot]\right] + \tfrac{1}{3}\left[\tfrac{1}{1}\left[\tfrac{1}{1}[RL]\right]\right] \\
&\quad (\mu_2) && \tfrac{1}{3}[\bot] + \tfrac{1}{3}\left[\tfrac{1}{1}[ML]\right] + \tfrac{1}{3}\left[\tfrac{1}{1}[RL]\right] \\
&(\mu_1) && \tfrac{1}{2}[ML] + \tfrac{1}{2}[RL] \\
&(\pi) && \tfrac{1}{2}[M] + \tfrac{1}{2}[R]\,.
\end{aligned}
$$

An interpretation of this other solution is that *(iii)* we *intervene* to force the host to open the left door (L), and the game show halts otherwise. Because it is forced, the host's decision stops carrying any inferential information: we are equally likely to see it no matter where the prize is.

Monty Hall is a famously controversial problem [53, 69]. We claim that this controversy boils down to a difference in parenthesization: *when to normalize* does matter and — when parentheses normalize — parentheses matter. Monty Hall is a counterexample to associativity.

```
(do (prize door) ⟵ (do                    ≠   (do (prize) ⟵ uniformDoor
        (prize) ⟵ uniformDoor                     (door) ⟵ (do
        (door) ⟵ (host prize)                         (door) ⟵ (host prize)
        return (prize door))                          () ⟵ (observe L door)
    () ⟵ (observe L door)                             return (door))
    return (prize))                               return (door))
```

Fig. 1. Two interpretations of the Monty Hall problem as programs in the magmadic metalanguage.

Indeed, using the *magmadic metalanguage* we later introduce, the two Monty Hall variants become the two programs on Figure 1, where *(i)* $\mathsf{uniformDoor} = \frac{1}{3}[L] + \frac{1}{3}[M] + \frac{1}{3}[R]$; *(ii)* $(\mathsf{host}\ L) = \frac{1}{1}[R]$; and $(\mathsf{host}\ M) = \frac{1}{2}[R] + \frac{1}{2}[L]$; and $(\mathsf{host}\ R) = \frac{1}{1}[L]$; *(iii)* $(\mathsf{observe}\ x\ y) = \top$ when $x = y$ and $\bot$ otherwise — formally, $\top = \frac{1}{1}[]$ and $\bot = 0$. In particular, $(\mathsf{observe}\ L\ L) = \top$ and $(\mathsf{observe}\ L\ R) = \bot$.

1.2 Example — The Unclear Test problem

Does this transfer to numerical disagreements? Let us explicitly solve a standard *Test problem* (c.f. [31, §2]). Consider an illness *(i)* with a 33% prevalence, $\mathsf{prevalence} = \frac{1}{3}[I] + \frac{2}{3}[H]$, for which we test *(ii)* with 75% sensitivity, $(\mathsf{test}\ I) = \frac{3}{4}[P] + \frac{1}{4}[N]$, and 50% specificity, $(\mathsf{test}\ H) = \frac{1}{2}[P] + \frac{1}{2}[N]$. What is the posterior probability of having this illness after *(iii)* a positive test?

$$
\begin{aligned}
&(i) && \tfrac{1}{3}[I] + \tfrac{2}{3}[H] \\
&(ii) && \tfrac{1}{3}\left[\tfrac{3}{4}[IP] + \tfrac{1}{4}[IN]\right] + \tfrac{2}{3}\left[\tfrac{1}{2}[HP] + \tfrac{1}{2}[HN]\right] \\
&(\mu_1) && \tfrac{1}{4}[IP] + \tfrac{1}{12}[IN] + \tfrac{1}{3}[HP] + \tfrac{1}{3}[HN] \\
&(iii) && \tfrac{1}{4}\left[\tfrac{1}{1}[IP]\right] + \tfrac{1}{12}[\bot] + \tfrac{1}{3}\left[\tfrac{1}{1}[HP]\right] + \tfrac{1}{3}[\bot] \\
&(\mu_2) && \tfrac{3}{7}[IP] + \tfrac{4}{7}[HP]\,. \\
&(\pi) && \tfrac{3}{7}[I] + \tfrac{4}{7}[H]\,.
\end{aligned}
$$

Answering $\frac{3}{7}[I]$ is hopefully uncontroversial, but a slight modification of the problem lends itself to two interpretations. Let us now consider an *unclear testing problem*. The statement is the same as before, but the test result is itself uncertain — e.g., we got an intermediate result that means positive 75% of the time — and thus, observing a positive test yields a distribution over the actual result

(uncertain = $\frac{3}{4}[P] + \frac{1}{4}[N]$). A first solution states all of the information *(i,ii,iii,iv)* and multiplies and normalizes after each step (μ_1, μ_2, μ_3).

$$
\begin{aligned}
&(i) && \tfrac{3}{4}[P] + \tfrac{1}{4}[N] \\
&(ii) && \tfrac{3}{4}\left[\tfrac{1}{3}[PI] + \tfrac{2}{3}[PH]\right] + \tfrac{1}{4}\left[\tfrac{1}{3}[NI] + \tfrac{2}{3}[NH]\right] \\
&(\mu_1) && \tfrac{1}{4}[PI] + \tfrac{1}{2}[PH] + \tfrac{1}{12}[NI] + \tfrac{1}{6}[NH] \\
&(iii) && \tfrac{1}{4}\left[\tfrac{3}{4}[PIP] + \tfrac{1}{4}[PIN]\right] + \tfrac{1}{2}\left[\tfrac{1}{2}[PHP] + \tfrac{1}{2}[PHN]\right] \\
&&& \quad + \tfrac{1}{12}\left[\tfrac{3}{4}[NIP] + \tfrac{1}{4}[NIN]\right] + \tfrac{1}{6}\left[\tfrac{1}{2}[NHP] + \tfrac{1}{2}[NHN]\right] \\
&(\mu_2) && \tfrac{3}{16}[PIP] + \tfrac{1}{16}[PIN] + \tfrac{1}{4}[PHP] + \tfrac{1}{4}[PHN] \\
&&& \quad + \tfrac{1}{16}[NIP] + \tfrac{1}{48}[NIN] + \tfrac{1}{12}[NHP] + \tfrac{1}{12}[NHN] \\
&(iv) && \tfrac{3}{16}\left[\tfrac{1}{1}[PIP]\right] + \tfrac{1}{16}[\bot] + \tfrac{1}{4}\left[\tfrac{1}{1}[PHP]\right] + \tfrac{1}{4}[\bot] \\
&&& \quad + \tfrac{1}{16}[\bot] + \tfrac{1}{48}\left[\tfrac{1}{1}[NIN]\right] + \tfrac{1}{12}[\bot] + \tfrac{1}{12}\left[\tfrac{1}{1}[NHN]\right] \\
&(\mu_3) && \tfrac{9}{26}[PIP] + \tfrac{6}{13}[PHP] + \tfrac{1}{26}[NIN] + \tfrac{2}{13}[NHN] \\
&(\pi) && \tfrac{5}{13}[I] + \tfrac{8}{13}[H]\,.
\end{aligned}
$$

However, a second solution treats the original *Test problem* as a subproblem *(ii,iii,iv)*, normalizing internally (μ_2, μ_3) before normalizing globally (μ_1). These two solutions are almost identical — a quick calculation could dismiss both as a probability of roughly 37% of having the illness. Yet, they yield different numbers, arguably answering two different questions.

$$
\begin{aligned}
&(i) && \tfrac{3}{4}[P] + \tfrac{1}{4}[N] \\
&\quad(ii) && \tfrac{3}{4}\left[\tfrac{1}{3}[PI] + \tfrac{2}{3}[PH]\right] + \tfrac{1}{4}\left[\tfrac{1}{3}[NI] + \tfrac{2}{3}[NH]\right] \\
&\quad(iii) && \tfrac{3}{4}\left[\tfrac{1}{3}\left[\tfrac{3}{4}[PIP] + \tfrac{1}{4}[PIN]\right] + \tfrac{2}{3}\left[\tfrac{1}{2}[PHP] + \tfrac{1}{2}[PHN]\right]\right] \\
&&& \quad + \tfrac{1}{4}\left[\tfrac{1}{3}\left[\tfrac{3}{4}[NIP] + \tfrac{1}{4}[NIN]\right] + \tfrac{2}{3}\left[\tfrac{1}{2}[NHP] + \tfrac{1}{2}[NHN]\right]\right] \\
&\quad(\mu_2) && \tfrac{3}{4}\left[\tfrac{1}{4}[PIP] + \tfrac{1}{12}[PIN] + \tfrac{1}{3}[PHP] + \tfrac{1}{3}[PHN]\right] \\
&&& \quad + \tfrac{1}{4}\left[\tfrac{1}{4}[NIP] + \tfrac{1}{12}[NIN] + \tfrac{1}{3}[NHP] + \tfrac{1}{3}[NHN]\right] \\
&\quad(iv) && \tfrac{3}{4}\left[\tfrac{1}{4}\left[\tfrac{1}{1}[PIP]\right] + \tfrac{1}{12}[\bot] + \tfrac{1}{3}\left[\tfrac{1}{1}[PHP]\right] + \tfrac{1}{3}[\bot]\right] \\
&&& \quad + \tfrac{1}{4}\left[\tfrac{1}{4}[\bot] + \tfrac{1}{12}\left[\tfrac{1}{1}[NIN]\right] + \tfrac{1}{3}[\bot] + \tfrac{1}{3}\left[\tfrac{1}{1}[NHN]\right]\right] \\
&\quad(\mu_3) && \tfrac{3}{4}\left[\tfrac{3}{7}[PIP] + \tfrac{4}{7}[PHP]\right] + \tfrac{1}{4}\left[\tfrac{1}{5}[NIN] + \tfrac{4}{5}[NHP]\right] \\
&(\mu_1) && \tfrac{9}{28}[PIP] + \tfrac{3}{7}[PHP] + \tfrac{1}{20}[NIN] + \tfrac{1}{5}[NHN] \\
&(\pi) && \tfrac{13}{35}[I] + \tfrac{22}{35}[H]\,.
\end{aligned}
$$

Again, the *unclear test problem* becomes a counterexample to associativity: under the previous definitions, the two programs in Figure 2 yield different results. The second solution — the one that does not simply normalize at the end — is possibly the more familiar one to statisticians. Jacobs has studied these two solutions as *Pearl's update* and *Jeffrey's update*, respectively [31]. We provide an interpretation for both by the end of the article.

```
(do (patient result outcome) ⟵ (do
        (result) ⟵ uncertainty
        (patient) ⟵ prevalence
        (outcome) ⟵ (test patient)
        return (patient result outcome))
    () ⟵ (observe result outcome)
    return (patient))
```

≠

```
(do (result) ⟵ uncertainty
    (patient outcome) ⟵ (do
        (patient) ⟵ prevalence
        (outcome) ⟵ (test patient)
        () ⟵ (observe result outcome)
        return (patient outcome))
    return (patient))
```

Fig. 2. Two interpretations of the *Unclear test problem* as programs in the magmadic metalanguage.

Remark 1.1 (The explicit normalize operator). In unnormalized probabilistic semantics, encoding these updates requires an explicit operator — usually called normalize (e.g., [61]). Arguably, this means paying the price of an unnormalized semantics to force composition to be associative, and then paying the price of an extra primitive operator to bypass that same associativity. Instead, if we accept that composition is not associative, we get to keep our semantics normalized while rendering the normalize operator redundant. We claim that this shift in perspective allows us to more naturally express phenomena of interest in probabilistic and causal inference.

1.3 Accepting non-associativity

Normalized composition is not associative: is there an associative and normalized probabilistic semantics? Normalized probabilistic semantics mixes two kinds of effect — stochasticity and partiality — that we encode with the distribution monad ($\mathbf{D}$) and the maybe monad ($\mathbf{M}$), respectively.

Definition 1.2 (Distribution monad)**.** A *distribution* over a set is a finite formal sum[2] over its elements whose coefficients are positive and add up to exactly 1. The (finitary) *distribution monad*, $\mathbf{D}\colon \mathrm{Set} \to \mathrm{Set}$, assigns, to each set, the set of distributions over it.

$$\mathbf{D}X = \left\{ \sum_{i=1}^{n} \lambda_i [x_i] \;\middle|\; x_i \in X, \lambda_i \in \mathbb{R}^+, \sum_{i=1}^{n} \lambda_i = 1 \right\}$$

Its *multiplication*, $\mu^{\mathbf{D}}\colon \mathbf{DD}X \to \mathbf{D}X$, is defined by $\mu^{\mathbf{D}}(\sum_i \lambda_i [\sum_j \mu_{ij} [x_{ij}]]) = \sum_i \sum_j \lambda_i \mu_{ij} [x_{ij}]$; its *unit*, $\eta^{\mathbf{D}}\colon X \to \mathbf{D}X$, is defined by $\eta^{\mathbf{D}}(x) = [x]$.

Definition 1.3 (Maybe monad)**.** The *maybe monad* (or, *option monad*), $\mathbf{M}\colon \mathrm{Set} \to \mathrm{Set}$, assigns to each set its *one-element extension* — with the extra disjoint element usually denoted by $\bot \in \mathbf{M}X$. That is, $\mathbf{M}X = \{\langle x \rangle \mid x \in X\} + \{\bot\}$. Its *multiplication*, $\mu^{\mathbf{M}}\colon \mathbf{MM}X \to \mathbf{M}X$, is defined by $\mu(\langle\langle x \rangle\rangle) = \langle x \rangle$ for $x \in X$, by $\mu(\langle \bot \rangle) = \bot$ and by $\mu(\bot) = \bot$; its *unit*, $\eta^{\mathbf{M}}\colon X \to \mathbf{M}X$, is defined by $\eta(x) = \langle x \rangle$.

Our examples mixed stochasticity and partiality in a particular way: either we had a full distribution or we had the empty distribution ($\bot$). *Could the following functor have a monad structure?*

$$\mathbf{D}_{\circ}X = \left\{ \sum_{i=1}^{n} \lambda_i [x_i] \;\middle|\; x \in X, \lambda_i \in \mathbb{R}^+, \sum_{i=1}^{n} \lambda_i = 1 \text{ or } \sum_{i=1}^{n} \lambda_i = 0 \right\} \cong \mathbf{MD}X.$$

Monads compose when there is a distributive law between them; and viceversa, under mild compatibility assumptions, a distributive law is needed to compose monads [2]. A distributive law of stochasticity over partiality is any consistent conversion of subdistributions into normalized distributions. For instance, a degenerate distributive law of stochasticity over partiality provides the so-called *black-hole* semantics [58, Example 3.1.(1)]. Black-hole semantics equates any non-full distribution to the empty distribution; it is the transformation $(-)^{\bot}\colon \mathbf{DM}X \to \mathbf{MD}X$ defined by

$$\left(\sum_{i=1}^{n} \lambda_i [\langle x_i \rangle] + \lambda_0 [\bot] \right)^{\bot} = \begin{cases} \left\langle \sum_{i=1}^{n} \lambda_i [x_i] \right\rangle, & \text{if } \sum_{i=0}^{n} \lambda_i = 1 \text{ and } \lambda_0 = 0, \\ \bot, & \text{otherwise.} \end{cases}$$

Alas, *black-hole semantics* is not helpful for our purposes: by equating any probability of failure to failure, it collapses to failure ($\bot$) any problem involving an uncertain observation. Worse still, *black-hole semantics* is the only valid distributive law.[3] No other monad distributes stochasticity over partiality; any non-trivial normalized semantics must drop some monad axiom.

[2] A finite formal sum is an element of $\mathbb{R}^+\langle X \rangle$, the free $\mathbb{R}^+$-module over the set X.

[3] This fact can also be derived from a classification of single-point extensions of convex algebras that appeared first in the work of Sokolova and Woracek [58]. The extended version keeps both a direct proof and an equivalence to this result [37].

Theorem (No-go result, Theorem 2.10)**.** *Black-hole semantics constitutes the only distributive law between the distribution monad and the maybe monad,* **DM → MD***.*

We drop associativity to introduce a novel and normalized-by-construction semantics for probabilistic and causal inference. Arguably, this semantics had been neglected because of its non-associativity: normalization forms a "sesquilaw", not a distributive law (Theorem 3.4); normalized distributions form a magmad, not a monad (Proposition 4.10, Theorem 4.5, and Corollary 4.6); normalized kernels form a Kleisli magmoid, not a Kleisli category (Proposition 4.10). We prove that associativity was unnecessary: we recover the usual unnormalized semantics by systematically left-associating (Theorem 4.5). We illustrate how non-associativity grants more expressiveness: it enables a unified magmadic metalanguage for both probabilistic and causal inference (Section 5).

1.4 Contributions

Normalization is not a distributive law (Theorem 2.9) even when it satisfies three of the axioms of distributive laws (Proposition 3.2). Moreover, there is no non-trivial distributive law of the distribution monad over the maybe monad (Theorem 2.10).

We introduce "sesquilaws" (Definition 3.3) — *one and a half distributive laws* that interact appropriately. We propose them as an axiomatization of the properties of normalization (Theorem 3.4) and we prove that normalization in standard Borel spaces and possibilistic post-selection are sesquilaws (Proposition 3.8 and Theorem 3.13); we relate these via sesquilaw morphisms and we prove that support — known to be a monad morphism (c.f. [20]) — is also a sesquilaw morphism (Propositions 3.17 and 3.19). In this abstract setting, we derive basic program simplifications (e.g., Lemma 3.5).

We then study "magmads" (Definition 4.1) and we prove that a sesquilaw induces a magmad, a monad, and an action of the latter on the former (Theorem 4.5); moreover, in a sesquilaw, some effects do always associate (Proposition 4.12). We introduce two magmadic metalanguages, depending on associativity (Definition 4.19); and we use them to characterize Bayes' update, Pearl's update, and Jeffrey's update (Propositions 4.22 and 4.24).

Finally, we introduce reasoning principles for the metalanguage: representing equality checks (Section 5.1), conditionals (Section 5.4), copyable and discardable effects (Section 5.3), and a variant of commutativity that arises only in the non-associative case (Proposition 5.7). We use these to prove the front-door and back-door criteria from causal inference (Theorems 5.17 and 5.18).

1.5 Related work

Probabilistic programming semantics employs subdistributions, starting with Kozen's [35] and Panangaden's substochastic variant of the Giry monad [23, 48]. Since these, both operational [36, 49, 50, 71] and denotational semantics [34, 43, 67] predominantly account for sample rejection with subdistributional [5, 15] or unnormalized semantics [8, 13, 14, 60], with normalization sporadically justifying program equations [62]. *Probabilistic programming languages* either normalize as a program transformation [47], or let their inference algorithms handle it [25, 65].

Categorical probability theory, in a line of work starting from Golubtsov, Cho and Jacobs, and Fritz, has abstracted stochastic kernels into *Markov categories* [6, 17, 24]. Further work has abstracted substochastic kernels and partial stochastic kernels into *partial Markov categories* [10] and *quasi-Markov categories* [18], respectively. In particular, string diagrammatic methods model causal interventions [51] as syntactic substitution [16, 19, 32]. At the same time, string diagrammatic axiomatizations of normalization have been proposed: we highlight those in terms of *normalization boxes* [33, 40] and *partial Markov categories* [10]. Simpson's *probability sheaves* constitute another approach to synthetic probability theory [56, 57], for which Stein recently proposed a comparison

[63]. Jacobs' *hypernormalization* [30] and Garner's *tricocycloids* [21] form an abstract approach to normalization that, however, does not address its non-associativity.

Magmads and failing distributive laws are infrequent: Munch-Maccagnoni [46] proposed a non-natural monad-comonad distributive law [41] and its Kleisli category to unify call-by-name and call-by-value. In probability, mass and chance interpretations determine a non-natural transformation $\mathbf{DD} \to \mathbf{DD}$ satisfying all of the other axioms of distributive laws [66]. Weak distributive laws, instead, [3, 22, 64] appear in imprecise probability [26, 39, 44]: there is no distributive law between non-determinism and probability [68], and this affects probabilistic trace semantics [7, 55, 59]. Bonchi, Sokolova, and Vignudelli [4, §3.4] distinguish possibilistic *may*, *must*, and *may-must* semantics [9]; we relate these to normalized semantics via support (cf. [20]): the possibilistic analogue of the normalized distributions magmad is the powerset monad.

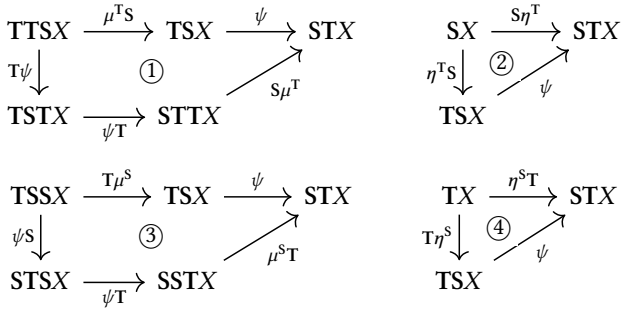
1.6 Synopsis

Section 2 proves that there is a single distributive law of the distribution monad over the maybe monad. Section 3 introduces sesquilaws, exemplified by both discrete and continuous probabilistic semantics and possibilistic semantics. Section 4 constructs magmads from sesquilaws, introduces the magmadic metalanguage, and introduces left-exchanging magmads. Section 4.5 derives Bayes' update, Pearl's update and Jeffrey's update from the magmadic structure. Section 5 picks some reasoning axioms and derives the front-door and back-door criteria.

2 Distributive Laws

Distributive laws (Definition 2.1) have well-known uses and limitations [2, 73]. Briefly, the composition of two monads is not a monad again, but distributive laws are sufficient — and necessary under mild conditions — to endow this composition with a monad structure (Proposition 2.2).

Definition 2.1 (Distributive law [2])**.** A *distributive law* of a monad, $(\mathbf{T}, \mu^{\mathbf{T}}, \eta^{\mathbf{T}})$, over a monad on the same category, $(\mathbf{S}, \mu^{\mathbf{S}}, \eta^{\mathbf{S}})$, is a natural transformation, $\psi \colon \mathbf{TS}X \to \mathbf{ST}X$, that makes the following four diagrams commute.



These four axioms of a distributive law are called ① $\mathbf{T}$-multiplicativity, ② $\mathbf{T}$-unitality, ③ $\mathbf{S}$-multiplicativity, and ④ $\mathbf{S}$-unitality, respectively.

Proposition 2.2 ([1, §9.2. Proposition 2.6])**.** *A distributive law, $\psi_X \colon \mathbf{TS}X \to \mathbf{ST}X$, between two monads, $(\mathbf{S}, \mu^{\mathbf{S}}, \eta^{\mathbf{S}})$ and $(\mathbf{T}, \mu^{\mathbf{T}}, \eta^{\mathbf{T}})$, is necessary and sufficient to induce a monad structure on the composite functor, $(\mathbf{ST}, \mu^{\mathbf{ST}}_{\psi}, \eta^{\mathbf{ST}}_{\psi})$, such that* (i) *the inclusions of the two monads, $\mathbf{S}\eta^{\mathbf{T}} \colon \mathbf{S}X \to \mathbf{ST}X$ and $\eta^{\mathbf{S}}\mathbf{T} \colon \mathbf{T}X \to \mathbf{ST}X$ are monad morphisms,* (ii) *the unit arises from tensoring, $\eta^{\mathbf{ST}}_{\psi} = \eta^{\mathbf{S}}\eta^{\mathbf{T}}$,* (iii) *and the* middle unitary equation *holds, $\mathbf{S}\eta^{\mathbf{T}}\eta^{\mathbf{S}}\mathbf{T} \fatsemi \mathbf{S}\psi\mathbf{T} \fatsemi \mu^{\mathbf{S}}\mu^{\mathbf{T}} = \mathbf{ST}$.*

This section first recalls the standard distributive law of subdistributional semantics: a distributive law of partiality over stochasticity (Section 2.1). The problem becomes to find a distributive law in

the other direction: a distributive law of stochasticity over partiality. We recall the formalization of *black-hole semantics* as a distributive law (Section 2.2). Finally, we prove first that normalization is not a distributive law (Theorem 2.9) and then the stronger result that there is no distributive law but *black-hole semantics* (Theorem 2.10).

Remark 2.3 (Monoidal monads). All of the monads we employ are monoidal monads: equivalently, they are *commutative strong* monads. All of the distributive laws we employ are moreover monoidal.

2.1 Example — Subdistributional semantics

Let us recall the distributive law of partiality over stochasticity that gives rise to subdistributional semantics (Proposition 2.5).[4] A *subdistribution* over a set is a finite formal sum over its elements whose coefficients are positive and add up to less or equal than 1. Equivalently, a subdistribution is a distribution over the one-element extension of the set: we use both interchangeably.

$$\mathbf{D}_{\leq 1}X = \left\{ \sum_{i=1}^{n} \lambda_i [x_i] \;\middle|\; x \in X, \lambda_i \in \mathbb{R}^+, \sum_{i=1}^{n} \lambda_i \leq 1 \right\} \cong \mathbf{DM}X.$$

Definition 2.4. The *subdistribution monad*, $(\mathbf{D}_{\leq 1}, \mu^{\mathbf{D}_{\leq 1}}, \eta^{\mathbf{D}_{\leq 1}})$, assigns, to each set, its subdistributions. Its *multiplication*, $\mu^{\mathbf{D}_{\leq 1}} : \mathbf{D}_{\leq 1}\mathbf{D}_{\leq 1}X \to \mathbf{D}_{\leq 1}X$, is given by $\mu^{\mathbf{D}_{\leq 1}}(\sum_i \lambda_i [\sum_j \mu_{ij}[x_{ij}]]) = \sum_i \sum_j \lambda_i \mu_{ij} [x_{ij}]$. Its *unit*, $\eta^{\mathbf{D}_{\leq 1}} : X \to \mathbf{D}_{\leq 1}X$, is given by $\eta^{\mathbf{D}_{\leq 1}}(x) = [x]$.

Proposition 2.5. *There exists a distributive law of the maybe monad over the distribution monad,* $(-)^\bullet : \mathbf{MD}X \to \mathbf{DM}X$*, defined by* $(\bot)^\bullet = \frac{1}{1}[\bot]$ *and* $\langle \sum_i \lambda_i [x_i] \rangle^\bullet = \sum_i \lambda_i [\langle x_i \rangle]$*. It induces a monad isomorphic to the subdistribution monad,* $(\mathbf{D}_{\leq 1}, \mu^{\mathbf{D}_{\leq 1}}, \eta^{\mathbf{D}_{\leq 1}}) \cong (\mathbf{DM}, \mu^{\mathbf{DM}}_\bullet, \eta^{\mathbf{DM}}_\bullet)$.

2.2 Example — Black-hole semantics

Perhaps surprisingly, there exists a distributive law of stochasticity over partiality [58] that gives rise to normalized distributions, albeit with black-hole semantics.[5] A *normalized distribution* over a set is a finite formal sum over its elements whose coefficients are positive and add up to 1 or 0. Equivalently, a normalized distribution is either a distribution or nothing: we use both interchangeably.

$$\mathbf{D}_\bot X = \left\{ \sum_{i=1}^{n} \lambda_i [x_i] \;\middle|\; x \in X, \lambda_i \in \mathbb{R}^+, \sum_{i=1}^{n} \lambda_i = 1 \text{ or } \sum_{i=1}^{n} \lambda_i = 0 \right\} \cong \mathbf{MD}X.$$

Definition 2.6. The *partial distribution monad*, $(\mathbf{D}_\bot, \mu^{\mathbf{D}_\bot}, \eta^{\mathbf{D}_\bot})$, assigns, to each set, its set of normalized distributions. Its *multiplication*, $\mu^{\mathbf{D}_\bot} : \mathbf{D}_\bot\mathbf{D}_\bot X \to \mathbf{D}_\bot X$, is defined by $\mu^{\mathbf{D}_\bot}(\sum_i \lambda_i [\sum_j \mu_{ij}[x_{ij}]]) = \sum_{ij} \lambda_i \mu_{ij} [x_i]$ whenever $\sum_{ij} \lambda_i \mu_{ij} = 1$, and by $\mu^{\mathbf{D}_\bot}(\sum_i \lambda_i [\sum_j \mu_{ij}[x_{ij}]]) = 0$ whenever $\sum_{ij} \lambda_i \mu_{ij} \neq 1$; its *unit*, $\eta^{\mathbf{D}_\bot} : X \to \mathbf{D}_\bot X$, is given by $\eta^{\mathbf{D}_\bot}(x) = [x]$.

Proposition 2.7. *There exists a distributive law of the distribution monad over the maybe monad,* $(-)^\bot : \mathbf{DM}X \to \mathbf{MD}X$*, defined by* $(\sum_{i=0}^n \lambda_i [\langle x_i \rangle])^\bot = \langle \sum_{i=0}^n \lambda_i [x_i] \rangle$ *when* $\sum_{i=0}^n \lambda_i = 1$*; and* $(\sum_{i=0}^n \lambda_i [x_i])^\bot = \bot$ *when* $\sum_{i=0}^n \lambda_i \neq 1$*. It induces a monad isomorphic to the partial distribution monad,* $(\mathbf{D}_\bot, \mu^{\mathbf{D}_\bot}, \eta^{\mathbf{D}_\bot}) \cong (\mathbf{MD}, \mu^{\mathbf{MD}}_\bot, \eta^{\mathbf{MD}}_\bot)$.

2.3 Counterexample — Normalized semantics

There is no other distributive law of stochasticity over partiality (Theorem 2.10). Normalization (Definition 2.8) satisfies all of the axioms of a distributive law except for one: the $\mathbf{D}$-multiplicativity axiom (Theorem 2.9).

[4] This is a particular case of Beck's distributive law of partiality over any effect [2, §4.2]

[5] Partial distributions and kernels are studied as quasi-Markov categories [54] but do not provide updating semantics.

Definition 2.8 (Normalization)**.** *Normalization*, $(-)^\circ \colon \mathbf{DM}X \to \mathbf{MD}X$, is the natural transformation defined by *(i)* division with the mass of a subdistribution whenever not zero, that *(ii)* fails otherwise.

$$(i) \quad \left(\textstyle\sum_{i=1}^n \lambda_i [\langle x_i \rangle] + \lambda_0 [\bot]\right)^\circ = \left\langle \sum_{i=1}^n \frac{\lambda_i}{\sum_{i=1}^n \lambda_i} [x_i] \right\rangle \quad \text{and} \quad (ii) \quad ([\bot])^\circ = \bot.$$

Theorem 2.9. *Normalization, $(-)^\circ \colon \mathbf{DM}X \to \mathbf{MD}X$, is not a distributive law of the distribution monad over the maybe monad.*

Proof. We prove that D-multiplicativity axiom fails. Consider the set $X = \{x, y\}$ and the following distribution of subdistributions: $\frac{1}{2}\left[\frac{1}{3}[\langle x\rangle] + \frac{2}{3}[\bot]\right] + \frac{1}{2}\left[\frac{2}{3}[\langle y\rangle] + \frac{1}{3}[\bot]\right] \in \mathbf{DDM}X$.

We will compute both sides of the D-multiplicativity axiom (①, Definition 2.1) and reach different expressions. On the one hand, we *(i)* normalize internally, *(ii)* normalize externally and *(iii)* multiply.

$$\begin{aligned} \mathbf{M}\mu^{\mathrm{D}} \circ \psi\mathbf{D} \circ \mathbf{D}\psi \left(\tfrac{1}{2}\left[\tfrac{1}{3}[\langle x\rangle] + \tfrac{2}{3}[\bot]\right] + \tfrac{1}{2}\left[\tfrac{2}{3}[\langle y\rangle] + \tfrac{1}{3}[\bot]\right]\right) &\overset{(i)}{=} \\ \mathbf{M}\mu^{\mathrm{D}} \circ \psi\mathbf{D} \left(\tfrac{1}{2}[\langle \tfrac{1}{1}[x]\rangle] + \tfrac{1}{2}[\langle \tfrac{1}{1}[y]\rangle]\right) &\overset{(ii)}{=} \\ \mathbf{M}\mu^{\mathrm{D}} \left(\langle \tfrac{1}{2}[[x]] + \tfrac{1}{2}[[y]]\rangle\right) &\overset{(iii)}{=} \\ \langle \tfrac{1}{2}[x] + \tfrac{1}{2}[y]\rangle\,. \end{aligned}$$

On the other hand, *(i)* we multiply and then *(ii)* normalize.

$$\begin{aligned} \psi \circ \mu^{\mathrm{D}}\mathbf{M} \left(\tfrac{1}{2}\left[\tfrac{1}{3}[\langle x\rangle] + \tfrac{2}{3}[\bot]\right] + \tfrac{1}{2}\left[\tfrac{2}{3}[\langle y\rangle] + \tfrac{1}{3}[\bot]\right]\right) &\overset{(i)}{=} \\ \psi\left(\tfrac{1}{6}[\langle x\rangle] + \tfrac{1}{3}[\langle y\rangle] + \tfrac{1}{2}[\bot]\right) &\overset{(ii)}{=} \\ \left\langle \tfrac{1}{3}[x] + \tfrac{2}{3}[y]\right\rangle\,. \end{aligned}$$

However, $\frac{1}{3}[x] + \frac{2}{3}[y] \neq \frac{1}{2}[x] + \frac{1}{2}[y]$, contradicting D-multiplicativity. □

Theorem 2.10 (No-go result, cf. [58, Theorem 5.3])**.** *Black-hole semantics constitutes the only distributive law between the distribution monad and the maybe monad,* $\mathbf{DM} \to \mathbf{MD}$.

Remark 2.11. Thus, normalization fails one of the axioms of distributive laws, but still satisfies most of the axioms and some further properties. The next section abstracts all of the structure of normalization into an alternative notion: that of sesquilaw.

3 Sesquilaws

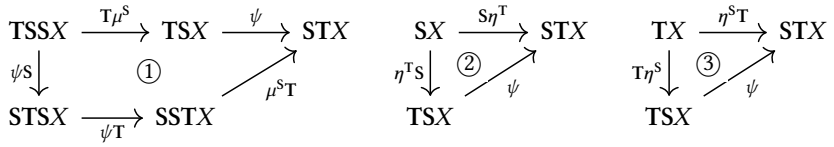
This section introduces sesquilaws: sections to an actual distributive law that satisfy all distributive law axioms except for the first one, which is only satisfied up-to-idempotent (Section 3.2). In particular, because they fail one of the axioms, they are quasidistributive laws; let us describe this more general concept first (Section 3.1).

3.1 Quasidistributive laws

A *quasidistributive law* could be any candidate distributive law failing one of the axioms. Specifically, we could define non-S-multiplicative, non-S-unital, non-T-multiplicative, and non-T-unital quasidistributive laws, respectively. In this terminology, a *weak distributive law* [3, 22, 26, 64] would be a non-T-unital quasidistributive law or, sometimes, a non-S-unital quasidistributive law. For the rest of the text, however, we focus on non-T-multiplicative quasidistributive laws, and simply call them quasidistributive laws.

Definition 3.1 (Quasidistributive law)**.** A (non-T-multiplicative) *quasidistributive law* of a monad, $(\mathbf{T}, \mu^{\mathrm{T}}, \eta^{\mathrm{T}})$, over a monad on the same category, $(\mathrm{S}, \mu^{\mathrm{S}}, \eta^{\mathrm{S}})$, is a natural transformation, $\psi \colon \mathbf{T}\mathrm{S}X \to$

$\mathrm{ST}X$, that makes the following three diagrams commute.



Proposition 3.2. *Normalization,* $(-)^\circ \colon \mathrm{DM}X \to \mathrm{MD}X$*, forms a quasidistributive law.*

Proof. For M-multiplicativity (①), we note that a generic element of $\mathrm{DMM}X$ must be of the form $\sum_i \lambda_i\,[x_i] + \mu_1\,[\bot_1] + \mu_2\,[\bot_2]$; let us reason by cases. If $\mu_1 + \mu_2 = 1$, then both sides of the equation yield $\bot$; either because $\mu_2 = 1$, or because, whenever $\mu_2 \neq 1$, it must be the case that $\mu_1 \neq 0$ and $\mu_1/\mu_1 = 1$. Let us assume that $\mu_1 + \mu_2 \neq 1$ and let us define $\lambda = \sum_i \lambda_i$. On the one side, we *(i)* use the definition of multiplication of the maybe monad, and *(ii)* the definition of normalization.

$$\begin{aligned} &\psi \circ \mathrm{D}\mu^{\mathrm{M}}\,(\textstyle\sum_i \lambda_i\,[\langle\langle x_i\rangle\rangle] + \mu_1\,[\langle\bot\rangle] + \mu_2\,[\bot]) && \overset{(i)}{=} \\ &\psi\,(\textstyle\sum_i \lambda_i\,[\langle x_i\rangle] + (\mu_1+\mu_2)\,[\bot]) && \overset{(ii)}{=} \\ &\langle\textstyle\sum_i \frac{\lambda_i}{\lambda}\,[x_i]\rangle. \end{aligned}$$

On the other side, we *(i)* consider the generic element, *(ii)* normalize at $\bot_2$, *(iii, iv)* normalize at $\bot_1$ and multiply using the maybe monad. By assumption, $\lambda \neq 0$ and $(\lambda_i/\lambda + \mu_1)(\lambda + \mu_1/\lambda) = \lambda_i/\lambda$.

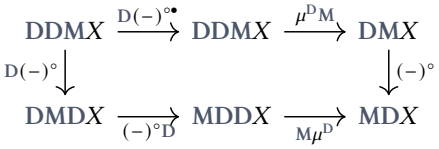
$$\begin{aligned} &\mu^{\mathrm{M}}\mathrm{D} \circ \mathrm{M}\psi \circ \psi\mathrm{M}\,(\textstyle\sum_i \lambda_i\,[\langle\langle x_i\rangle\rangle] + \mu_1\,[\langle\bot\rangle] + \mu_2\,[\bot]) && \overset{(i)}{=} \\ &\mu^{\mathrm{M}}\mathrm{D} \circ \mathrm{M}\psi\,(\langle\textstyle\sum_i \frac{\lambda_i}{\lambda+\mu_1}\,[\langle x_i\rangle] + \frac{\mu_1}{\lambda+\mu_1}\,[\bot]\rangle) && \overset{(ii)}{=} \\ &\mu^{\mathrm{M}}\mathrm{D}\,(\langle\langle\textstyle\sum_i \frac{\lambda_i}{\lambda}\,[x_i]\rangle\rangle) && \overset{(iii)}{=} \\ &\langle\textstyle\sum_i \frac{\lambda_i}{\lambda}\,[x_i]\rangle. \end{aligned}$$

For D-unitality (②), we reason by cases on $\mathrm{M}X$. We directly check that $(\eta^{\mathrm{D}}(\bot))^\circ = ([\bot])^\circ = \bot$ and $(\eta^{\mathrm{D}}(\langle x\rangle))^\circ = [\langle x\rangle]^\circ = \langle[x]\rangle$. Finally, for M-unitality (③), we must check that the normalization of a normalized distribution is itself: we note that, in that case, $\sum_i \lambda_i = 1$. □

Normalization, as a quasidistributive law, only induces a magmad. Together with the distributive law describing subdistributions, we have *"one and a half"* distributive laws that interact with each other: this is what we will call a "sesquilaw".

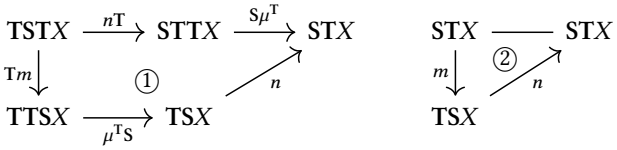
3.2 Sesquilaws

To recapitulate, normalization, $(-)^\circ \colon \mathrm{DM} \to \mathrm{MD}$, satisfies all distributive law axioms except for the D-multiplicativity axiom. Still, normalization satisfies a weaker version of this missing axiom: note that normalization has a section, $(-)^\bullet \colon \mathrm{MD} \to \mathrm{DM}$, and it induces an idempotent, $(-)^{\circ\bullet} = (-)^{\circ\bullet\circ\bullet}$; multiplicativity holds when we precompose one side with this idempotent.



This section introduces sesquilaws, which abstract this setting into two commutative diagrams (Definition 3.3). Normalization is a sesquilaw (Theorem 3.4); we will use the sesquilaw axioms recover the properties of normalization (Lemma 3.5, Theorem 4.5, and Propositions 4.12 and 5.8).

Definition 3.3 (Sesquilaw)**.** A *sesquilaw*, $(\mathbf{S}, \mathbf{T}, m, n)$, of a monad $(\mathbf{T}, \mu^{\mathbf{T}}, \eta^{\mathbf{T}})$ over a monad on the same category $(\mathbf{S}, \mu^{\mathbf{S}}, \eta^{\mathbf{S}})$ consists of a distributive law, $m\colon \mathbf{ST}X \to \mathbf{TS}X$, and a quasidistributive law, $n\colon \mathbf{TS}X \to \mathbf{ST}X$, additionally making the following two diagrams commute.



Theorem 3.4. *Normalized semantics,* $(-)^{\circ}\colon \mathbf{DM}X \to \mathbf{MD}X$*, and subdistributional semantics,* $(-)^{\bullet}\colon \mathbf{MD}X \to \mathbf{DM}X$*, form a sesquilaw of the distribution monad over the maybe monad.*

Proof. We already know that normalization and the inclusion into subdistributions are one-sided inverses (axiom ②), and that normalization forms a quasidistributive law (Proposition 3.2). Let us prove that the first diagram commute (axiom ①).



Let us note that a generic element of $\mathbf{DMD}X$ is, equivalently, a subdistribution of distributions, $\sum_i \lambda_i \left[\left\langle \sum_j \mu_{ij} \left[x_{ij}\right]\right\rangle\right] + \lambda_0[\bot]$, where the definition ensures $\sum_j \mu_{ij} = 1$ for each $i = 0, ..., n$. Let us write $\lambda = \sum_i \lambda_i$; whenever $\lambda = 0$, the desired equation automatically holds. Let us first *(i)* normalize and *(ii)* flatten the distributions.

$$\begin{aligned} &\mathbf{M}\mu \circ n\mathbf{D}\left(\textstyle\sum_i \lambda_i \left[\left\langle \sum_j \mu_{ij} \left[x_{ij}\right]\right\rangle\right] + \lambda_0[\bot]\right) && \overset{(i)}{=} \\ &\quad \mathbf{M}\mu\left(\left\langle \textstyle\sum_i \frac{\lambda_i}{\lambda} \left[\sum_j \mu_{ij} \left[x_{ij}\right]\right]\right\rangle\right) && \overset{(ii)}{=} \\ &\quad \left\langle \textstyle\sum_{ij} \frac{\lambda_i \mu_{ij}}{\lambda} \left[x_{ij}\right]\right\rangle. \end{aligned}$$

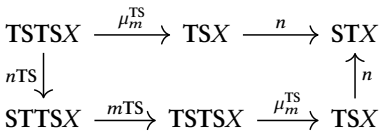
On the other hand, let us *(i)* regard the potential distributions as subdistributions, via subdistributional semantics, *(ii)* flatten the distributions, and *(iii)* normalize.

$$\begin{aligned} &n \circ \mu\mathbf{M} \circ \mathbf{D}m\left(\textstyle\sum_i \lambda_i \left[\left\langle \sum_j \mu_{ij} \left[x_{ij}\right]\right\rangle\right] + \lambda_0[\bot]\right) && \overset{(i)}{=} \\ &\quad n \circ \mu\mathbf{M}\left(\textstyle\sum_i \lambda_i \left[\sum_j \mu_{ij} \left[\left\langle x_{ij}\right\rangle\right]\right] + \lambda_0[[\bot]]\right) && \overset{(ii)}{=} \\ &\quad n\left(\textstyle\sum_{ij} \lambda_i \mu_{ij} \left[\left\langle x_{ij}\right\rangle\right] + \lambda_0[\bot]\right) && \overset{(iii)}{=} \\ &\quad \left\langle \textstyle\sum_{ij} \frac{\lambda_i \mu_{ij}}{\sum_{ij} \lambda_i \mu_{ij}} \left[x_{ij}\right]\right\rangle. \end{aligned}$$

The last step is to see that $\sum_{ij} \lambda_i \mu_{ij} = \lambda$, which follows from $\sum_j \mu_{ij} = 1$ for each $i = 0, ..., n$. □

Sesquilaws contain enough structure to abstractly prove useful facts about normalization. For instance, the following rule — which we call *renormalization* (Lemma 3.5) — states that normalizing after each multiplication is equivalent to normalizing at the end. It will serve us later to justify the soundness of normalizing and resampling after each transformation (Theorem 4.5).

Lemma 3.5 (Renormalization)**.** *Any sesquilaw,* $(\mathbf{S}, \mathbf{T}, m, n)$*, induces an idempotent,* $k = (n \fatsemi m)\colon \mathbf{TS} \to \mathbf{TS}$*. This idempotent is left-absorptive, meaning that the following diagram commutes.*



Proof sketch. The proof uses naturality, the definition of the monad arising from a distributive law, the S-multiplicativity of the quasidistributive law n, the first sesquilaw axiom, and the second sesquilaw axiom. See the extended version for details [37]. □

3.3 Example — Normalized boolean semantics is associative

Possibilistic semantics is encoded by the affine powerset monad (Definition 3.6), which distributes over the maybe monad with a sesquilaw (Proposition 3.8). This sesquilaw is a simple form of normalization — removing the possibility of failure — that we call *postselection* (Definition 3.7).

Definition 3.6 (Affine powerset monad)**.** The *affine powerset monad* (or, *non-empty finitary powerset monad*), $\mathbf{R}\colon \mathrm{Set} \to \mathrm{Set}$, assigns, to each set, its set of non-empty finite subsets,

$$\mathbf{R}X = \{U \subseteq X \mid U \text{ finite, and } U \neq \varnothing\}.$$

Its *multiplication*, $\mu^{\mathrm{R}}\colon \mathbf{RR}X \to \mathbf{R}X$, is defined by $\mu^{\mathrm{R}}(A) = \bigcup_{X \in A} X$; its *unit*, $\eta^{\mathrm{R}}\colon X \to \mathbf{R}X$, is defined by $\eta^{\mathrm{R}}(x) = \{x\}$.

Definition 3.7 (Postselection)**.** Postselection, $\mathrm{p}(-)\colon \mathbf{RM}X \to \mathbf{MR}X$, is the natural transformation that removes failure if it can and fails only if it must. It is defined by

$$\mathrm{p}(A \cup \{\bot\}) = \mathrm{p}(A) = \langle A \rangle\,, \text{ for each } A \neq \{\bot\}, \text{ and } \mathrm{p}(\{\bot\}) = \bot.$$

Postselection is a retraction of the inclusion $\mathrm{i}\colon \mathbf{MR}X \to \mathbf{RM}X$ defined by $\mathrm{i}(\bot) = \{\bot\}$ and $\mathrm{i}(\langle A \rangle) = \{\langle a \rangle \mid a \in A\}$ for each $A \in \mathbf{R}X$.

Proposition 3.8 (Normalized boolean semantics)**.** *Postselection and inclusion form a sesquilaw,* $(\mathbf{M}, \mathbf{R}, \mathrm{i}, \mathrm{p})$*, of the affine powerset monad over the maybe monad.*

Remark 3.9. The magmad associated to the sesquilaw, $(\mathbf{MR}, \mu_{\mathrm{p}}^{\mathbf{MR}}, \eta_{\mathrm{p}}^{\mathbf{MR}})$ happens to be a monad — it is equivalent to the finite powerset monad. The monad associated to the sesquilaw, $(\mathbf{RM}, \mu_{\mathrm{i}}^{\mathbf{RM}}, \eta_{\mathrm{i}}^{\mathbf{RM}})$, is the monad of subaffine relations used for *may-must semantics* (e.g., [4, §3.4]). Moreover, there exists another sesquilaw determining possibilistic *black-hole semantics* [4, 44].

3.4 Example — Normalized continuous probabilistic semantics is not associative

Sesquilaws are abstract enough to unify the discrete and the continuous case; this section details normalization in standard Borel spaces. We recall the Giry monad (Definition 3.10) and Panangaden's substochastic variant (Proposition 3.11).

Definition 3.10 (Giry monad, [23, 38])**.** The functor taking a Borel space to the set of probability measures over it, $\mathcal{D}\colon \mathbf{BorelMes} \to \mathbf{BorelMes}$, forms a monad, $(\mathcal{D}, \eta^{\mathcal{D}}, \mu^{\mathcal{D}})$, whose unit, $\eta^{\mathcal{D}}\colon X \to \mathcal{D}X$, is the Dirac delta, $\eta^{\mathcal{D}}(x) = \delta(x; -)$, and whose multiplication, $\mu^{\mathcal{D}}\colon \mathcal{D}\mathcal{D}X \to \mathcal{D}X$, is given by Lebesgue integration,

$$\mu^{\mathcal{D}}(\Omega)(U) = \int_{\nu \in \mathcal{D}X} \nu(U) \cdot \Omega(\mathrm{d}\nu).$$

Proposition 3.11 (Panangaden monad, [48, §3])**.** *Standard Borel spaces admit coproducts and thus a Maybe monad,* $\mathcal{M}\colon \mathbf{BorelMes} \to \mathbf{BorelMes}$*. There exists a distributive law,* $(-)^{\bullet}_{X}\colon \mathcal{M}\mathcal{D}X \to \mathcal{D}\mathcal{M}X$*, yielding a monad structure on* $(\mathcal{D}\mathcal{M}, \mu_{\bullet}^{\mathcal{D}\mathcal{M}}, \eta_{\bullet}^{\mathcal{D}\mathcal{M}})$*.*

Definition 3.12 (Continuous normalization)**.** *Normalization*, $\mathrm{N}(-)\colon \mathcal{D}\mathcal{M}X \to \mathcal{M}\mathcal{D}X$, is the natural transformation defined by *(i)* division with the mass of a subdistribution whenever not zero, and *(ii)* zero otherwise.

$$(i)\quad \mathrm{N}\nu(U) = \frac{\nu(U)}{\nu(X)}, \text{ when } \nu(X) \neq 0, \text{ and} \qquad (ii)\quad \mathrm{N}\nu(U) = \bot \text{ when } \nu(X) = 0.$$

Theorem 3.13. *Normalization in standard Borel spaces,* $\mathsf{N}(-)\colon \mathcal{D}\mathcal{M}X \to \mathcal{M}\mathcal{D}X$*, yields a sesquilaw.*

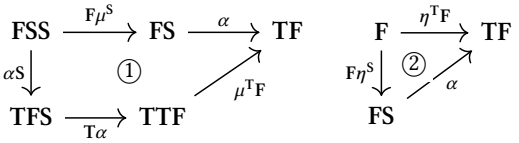
Proof sketch. Monoidality follows from linearity of the Lebesgue integral. The sesquilaw axiom follows ultimately from the monotone convergence theorem, employed as for the associativity of the Giry monad [48]. See the extended version for details [37]. □

3.5 Sesquilaw morphisms

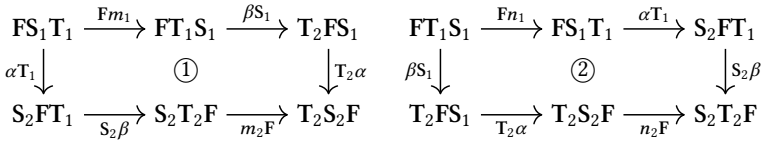
All these notions of normalization — discrete, possibilistic, and continuous — form sesquilaws; we can translate between these sesquilaws via morphisms. This section introduces sesquilaw morphisms (Definition 3.15). We start by recalling monad morphisms (Definition 3.14) and we present two examples of these translations at the end of the section.

It may be more natural for this section to use double-categorical pasting diagrams, but we present it using standard commutative diagrams for consistency.

Definition 3.14 (Monad morphism)**.** A *monad morphism* from a monad $(\mathbf{S}, \mu^{\mathbf{S}}, \eta^{\mathbf{S}})$ on a category $\mathbb{A}_1$ to another monad $(\mathbf{T}, \mu^{\mathbf{T}}, \eta^{\mathbf{T}})$ on a category $\mathbb{A}_2$ consists of a functor $\mathbf{F}\colon \mathbb{A}_1 \to \mathbb{A}_2$ and a natural transformation $\alpha\colon \mathbf{FS} \to \mathbf{TF}$ such that the following two diagrams commute.



Definition 3.15 (Sesquilaw morphism)**.** A *sesquilaw morphism* from a sesquilaw $(\mathbf{S}_1, \mathbf{T}_1, m_1, n_1)$ on a category $\mathbb{A}_1$ to another sesquilaw $(\mathbf{S}_2, \mathbf{T}_2, m_2, n_2)$ on a category $\mathbb{A}_2$ is a pair of monad morphisms $\alpha\colon \mathbf{FS}_1 \to \mathbf{S}_2\mathbf{F}$ and $\beta\colon \mathbf{FT}_1 \to \mathbf{T}_2\mathbf{F}$ such that the following two diagrams commute.



Example 3.16 (Support). The *support* of a distribution is the subset of elements present in its formal sum. Support determines a natural transformation, $\mathrm{supp}_X\colon \mathbf{D}X \to \mathbf{R}X$, defined by

$$\mathrm{supp}\left(\sum_{i=1}^{n} \lambda_i [x_i]\right) = \{x_i \in X \mid \lambda_i \neq 0\}.$$

It is well-known that support is a monad morphism. We note that it is, moreover, a sesquilaw morphism: intuitively, the support of a normalization is the postselection of its support.

Proposition 3.17 (Support)**.** *Support,* $\mathrm{supp}_X\colon \mathbf{D}X \to \mathbf{R}X$*, is a sesquilaw morphism — with* $\mathbf{F} = \mathrm{Id}$ *the identity functor — between normalization and postselection.*

Example 3.18 (Discrete to continuous). Discrete normalization is a particular case of continuous normalization. Every set can be seen as a discrete standard Borel space, $\mathbf{F}\colon \mathbf{Set} \to \mathbf{BorelMes}$; and there exist monad morphisms $\alpha\colon \mathbf{FM}X \to \mathcal{M}\mathbf{F}X$ and $\beta\colon \mathbf{FD}X \to \mathcal{D}\mathbf{F}X$ regarding each finitary distribution as a probability measure.

Proposition 3.19 (Finitary distributions as probability measures)**.** *Inclusion of finitary distributions into probability measures,* $\beta\colon \mathbf{FD}X \to \mathcal{D}\mathbf{F}X$*, and the inclusion of the maybe monad,* $\alpha\colon \mathbf{FM}X \to \mathcal{M}\mathbf{F}X$*, form a sesquilaw morphism — with* $\mathbf{F}$ *the inclusion of sets into standard Borel spaces — between discrete and continuous normalization.*

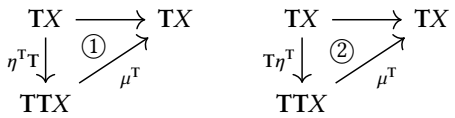
4 Magmads and magmoids

We now introduce magmads (Section 4.1) — *"non-associative monads"* — and develop our main example: the magmad of normalized distributions. Each magmad induces a Kleisli magmoid (Section 4.2) — a *"non-associative category"* — and enables two *magmadic metalanguages*: one associating to the right, and one associating to the left (Section 4.4). This section uses the sesquilaw axioms to reason about these metalanguages; it then applies the metalanguages to Bayes' update, Pearl's update, and Jeffrey's update (Propositions 4.22 and 4.24).

4.1 Magmads

A magmad is a monad failing the associativity axiom[6] (Definition 4.1). Magmads are infrequent in the literature. Particularly in functional programming, it is assumed that a monad needs to obey all of its axioms to be of interest. Still, quasidistributive laws only induce magmads (Proposition 4.2) — and normalized distributions only form a magmad.

Definition 4.1 (Magmad)**.** A *magmad*, $(\mathbf{T}, \mu^{\mathbf{T}}, \eta^{\mathbf{T}})$, consists of an endofunctor, $\mathbf{T}\colon \mathbb{C} \to \mathbb{C}$, together with two natural transformations — the unit, $\eta_X\colon X \to \mathbf{T}X$, and the multiplication, $\mu_X\colon \mathbf{TT}X \to \mathbf{T}X$ — making the following two diagrams commute.



Proposition 4.2. *Any quasidistributive law, $\psi\colon \mathbf{TS}X \to \mathbf{ST}X$, induces a magmad, $(\mathbf{ST}, \mu^{\mathbf{ST}}_{\psi}, \eta^{\mathbf{ST}}_{\psi})$ defined by $\mu^{\mathbf{ST}}_{\psi} = \mathbf{S}\psi\mathbf{T} \fatsemi \mu^{\mathbf{S}}\mu^{\mathbf{T}}$ and $\eta^{\mathbf{ST}}_{\psi} = \eta^{\mathbf{S}}\eta^{\mathbf{T}}$.*

Example 4.3 (Normalized distribution magmad). The *normalized distribution magmad*, $(\mathbf{D}_\circ, \mu^{\mathbf{D}_\circ}, \eta^{\mathbf{D}_\circ})$, assigns, to each set, the set of normalized distributions over it.[7]

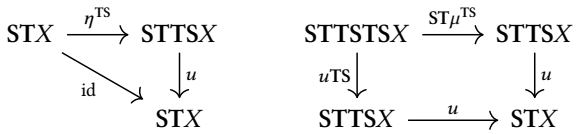
$$\mathbf{D}_\circ X = \left\{ \sum_{i=1}^{n} \lambda_i [x_i] \;\middle|\; x \in X, \lambda_i \in \mathbb{R}^+, \sum_{i=1}^{n} \lambda_i = 1 \text{ or } \sum_{i=1}^{n} \lambda_i = 0 \right\} \cong \mathbf{M}\mathbf{D}X.$$

Its multiplication, $\mu^{\mathbf{D}_\circ}\colon \mathbf{D}_\circ\mathbf{D}_\circ X \to \mathbf{D}_\circ X$, is defined by $\mu^{\mathbf{D}_\circ}(\sum_i \lambda_i\, [\sum_j \rho_{ij}[x_{ij}]]) = \sum_{ij} \frac{\lambda_i\rho_{ij}}{\sum_{ij}\lambda_i\rho_{ij}} [x_{ij}]$, which is zero whenever $\sum_{ij}\lambda_i\rho_{ij} = 0$.[8] Its *unit*, $\eta^{\mathbf{D}_\circ}\colon X \to \mathbf{D}_\circ X$, is defined by $\eta^{\mathbf{D}_\circ}(x) = [x]$.

Proposition 4.4. *The normalization sesquilaw (Theorem 3.4) induces a magmad (by Proposition 4.2) isomorphic to the normalized distribution magmad, $(\mathbf{D}_\circ, \mu^{\mathbf{D}_\circ}, \eta^{\mathbf{D}_\circ}) \cong (\mathbf{MD}, \mu^{\mathbf{MD}}_\circ, \eta^{\mathbf{MD}}_\circ)$.*

When the magmad arises from a sesquilaw it is moreover endowed with an action from the monad arising from the sesquilaw (Theorem 4.5). In our main example, subdistributions act on normalized distributions akin to how updates act on priors.

Theorem 4.5. *Any sesquilaw, $(\mathbf{S}, \mathbf{T}, m, n)$, induces a right action of its monad, $\mathbf{TS}$, on its magmad, $\mathbf{ST}$: i.e., a natural transformation, $u\colon \mathbf{STTS}X \to \mathbf{ST}X$, making the following two diagrams commute.*



[6] For this text, we adopt the convention that magmas — and any of their derived notions — are unital.
[7] This is the exact same functor and unit as in black-hole semantics (Section 2.2); the only difference is the multiplication.
[8] Whenever $\sum_{ij}\lambda_i\rho_{ij} = 0$, there are no formal summands and no division by zero: the result is the empty formal sum.

This action is defined by either side of the following commutative diagram.

$$\begin{array}{ccccc} \mathrm{STTS}X & \xrightarrow{\mathrm{S}\mu^{\mathrm{T}}\mathrm{S}} & \mathrm{STS}X & \xrightarrow{m\mathrm{S}} & \mathrm{TSS}X \\ {\scriptstyle m\mathrm{TS}}\downarrow & & & & \downarrow{\scriptstyle \mathrm{T}\mu^{\mathrm{S}}} \\ \mathrm{TSTS}X & \xrightarrow{\mathrm{T}m\mathrm{S}} & \mathrm{TTSS}X & \xrightarrow{\mu^{\mathrm{T}}\mu^{\mathrm{S}}} & \mathrm{TS}X \xrightarrow{n} \mathrm{ST}X \end{array}$$

Proof sketch. Multiplicativity follows from the two axioms of sesquilaws, together with **S**-multiplicativity of the quasidistributive law and the **T**-multiplicativity of the distributive law; we also use the associativity of both monads. Unitality follows from the unitality of both monads and the second axiom of sesquilaws. □

Corollary 4.6. *The subdistribution monad acts on the normalized distribution magmad: there exists a natural transformation* $u\colon \mathbf{D}_\circ\mathbf{D}_{\leq 1}X \to \mathbf{D}_\circ X$ *determining a right monad action.*

4.2 Kleisli Magmoids

Magmoids are *"non-associative categories"*: we repeat the definition of category without associativity (Definition 4.7, cf. [41]); all magmoids in this text are unital. Analogously to monads inducing Kleisli categories, magmads induce Kleisli magmoids (Proposition 4.10). Our main non-associative example is the magmoid of normalized stochastic kernels (Proposition 4.8): the Kleisli magmoid of the normalized distribution magmad.

Definition 4.7 (Magmoid)**.** A (unital) *magmoid*, $\mathbb{M}$, consists of morphisms, $\mathbb{M}(X;Y)$, between some collection of objects $X, Y \in \mathbb{M}_{obj}$, equipped with a composition operation, $(\fatsemi)\colon \mathbb{M}(X;Y) \times \mathbb{M}(Y;Z) \to \mathbb{M}(X;Z)$, and identities, $\mathrm{id} \in \mathbb{M}(X;X)$, that are unital, meaning that $f \fatsemi \mathrm{id} = f = \mathrm{id} \fatsemi f$.

Proposition 4.8. Normalized stochastic kernels *between sets,* $X \to \mathbf{M}\mathbf{D}Y$*, form a magmoid,* **Norm**. *Given* $f\colon X \to \mathbf{M}\mathbf{D}Y$*, we write* $f(x;y)$ *for the coefficient of* $y \in Y$ *in the normalized distribution* $f(x)$. *Composition of two morphisms,* $f\colon X \to \mathbf{M}\mathbf{D}Y$ *and* $g\colon Y \to \mathbf{M}\mathbf{D}Z$*, is defined by*

$$(f \fatsemi g)(x;z) = \frac{\sum_{v\in Y} f(x;v)\cdot g(v;z)}{\sum_{v\in Y}\sum_{w\in Z} f(x;v)\cdot g(v;w)}, \text{ when } \sum_{v,w} f(x;v)\cdot g(v;w) \neq 0; \text{ and } 0 \text{ otherwise.}$$

Remark 4.9. In other words, if we consider the associated substochastic kernels, $f^\bullet\colon X \to \mathbf{D}\mathbf{M}Y$ and $g^\bullet\colon Y \to \mathbf{D}\mathbf{M}Z$, it is the normalization of their composition, $f \fatsemi g = (f^\bullet ; g^\bullet)^\circ$.

Proposition 4.10 (Kleisli magmoids)**.** *Any magmad,* $(\mathbf{N}, \mu^{\mathbf{N}}, \eta^{\mathbf{N}})$ *over a category* $\mathbb{C}$ *with composition operation* $(;)$*, induces a magmoid,* $\mathbb{K}(\mathbf{N})$*, with its same objects, with morphisms* $\mathbb{K}(\mathbf{N})(X;Y) = \mathbb{C}(X;\mathbf{N}Y)$*, with identities* $\mathrm{id} = \eta^{\mathbf{N}}$*, and with compositions,* $(f \fatsemi g) = f ; \mathbf{N}g ; \mu^{\mathbf{N}}$.

Proposition 4.11. *The magmoid of normalized stochastic kernels (Proposition 4.8) is isomorphic to the Kleisli magmoid of the normalized distribution magmad,* $\mathbf{Norm} \cong \mathbb{K}(\mathbf{D}_\circ)$.

Associativity may still hold in some cases of interest. For instance, because the magmoid of normalized stochastic kernels arises from a sesquilaw, three morphisms commute whenever the first or the second are partial (of the form $X \to \mathbf{M}Y$) or the third is stochastic (of the form $X \to \mathbf{D}Y$).

Proposition 4.12 (Sesquilaw magmoids)**.** *Any sesquilaw,* $(\mathbf{S}, \mathbf{T}, m, n)$*, induces a Kleisli magmoid where associativity is satisfied,* $f \fatsemi (g \fatsemi h) = (f \fatsemi g) \fatsemi h$*, if either*

(1) $f = u ; \mathbf{S}\eta^{\mathbf{T}}$ *for some* $u\colon X \to \mathbf{S}Y$*;*
(2) $g = v ; \mathbf{S}\eta^{\mathbf{T}}$ *for some* $v\colon Y \to \mathbf{S}Z$*; or*
(3) $h = w ; \eta^{\mathbf{S}}\mathbf{T}$ *for some* $w\colon Z \to \mathbf{T}W$.

4.3 Examples — Monty Hall problem and Unclear Test problem

The magmoid of normalized stochastic kernels is not a category: let us recover our two original examples (Section 1). First of all, note that we have the ability to rearrange inputs according to any finite function $\sigma\colon \underline{m} \to \underline{n}$, thanks to cartesianity and strength (see Remark 2.3, cf. [42]).

Proposition 4.13. *Any strong magmad over a cartesian monoidal category induces a copying operator* $(-)_\star\colon \mathbb{K}(\mathbf{T})(X;Y) \to \mathbb{K}(\mathbf{T})(X;X\times Y)$*; moreover, each finite function* $\sigma\colon \underline{m}\to\underline{n}$ *induces*

(1) *a projection,* $\pi_\sigma\colon X_1\times\ldots\times X_n \to \mathbf{T}(X_{\sigma(1)}\times\ldots\times X_{\sigma(m)})$*;*
(2) *a rearrangement,* $(-)_\sigma\colon \mathbb{K}(\mathbf{T})(X_{\sigma(1)}\times\ldots\times X_{\sigma(m)};Y)\to\mathbb{K}(\mathbf{T})(X_1\times\ldots\times X_n;Y)$*;*
(3) *and rearrangement followed by copying of inputs,*

$$(-)_{\sigma^\star}\colon \mathbb{K}(\mathbf{T})(X_{\sigma(1)}\times\ldots\times X_{\sigma(m)};Y)\to\mathbb{K}(\mathbf{T})(X_1\times\ldots\times X_n;X_1\times\ldots\times X_n\times Y).$$

Example 4.14. The *Monty Hall* problem defines a set of doors, $X=\{L,M,R\}$, and a set of announcements, $Y=\{L,M,R\}$. We now define $f^1\colon 1\to \mathbf{D}_\circ X$ by $f^1=\frac{1}{3}[L]+\frac{1}{3}[M]+\frac{1}{3}[R]$; we define $f^2\colon X\to\mathbf{D}_\circ Y$ by $f^2(L)=\frac{1}{1}[R]$, and $f^2(M)=\frac{1}{2}[L]+\frac{1}{2}[R]$, and $f^2(R)=\frac{1}{1}[L]$; we define $g^3\colon Y\times Y\to\mathbf{D}_\circ(1)$ by $g^3(y_1,y_2)=\top$ whenever $y_1=y_2$ and $g^3(y_1,y_2)=\bot$ whenever $y_1\neq y_2$; we define $f^3\colon Y\to\mathbf{D}_\circ(1)$ as the partial application $f^3(y_2)=g^3(L,y_2)$. Rearranging inputs, we derive functions $f^2_{\sigma_2^\star}\colon X\to\mathbf{D}_\circ(X\times Y)$ and $f^3_{\sigma_3^\star}\colon X\times Y\to\mathbf{D}_\circ(X\times Y)$.

The two solutions to the *Monty Hall* problem (Section 1) correspond, at each line, to a factor of the following compositions — plus the final projection, $\pi\colon X\times Y\to\mathbf{D}_\circ(X)$, which we omit.

$$(f^1\fatsemi f^2_{\sigma_2^\star})\fatsemi f^3_{\sigma_3^\star}=f^1\mathbin{;}\mathbf{D}_\circ f^2_{\sigma_2^\star}\mathbin{;}\mu\mathbin{;}\mathbf{D}_\circ f^3_{\sigma_3^\star}\mathbin{;}\mu\neq f^1\mathbin{;}\mathbf{D}_\circ(f^2_{\sigma_2^\star}\mathbin{;}\mathbf{D}_\circ f^3_{\sigma_3^\star}\mathbin{;}\mu)\mathbin{;}\mu=f^1\fatsemi(f^2_{\sigma_2^\star}\fatsemi f^3_{\sigma_3^\star}).$$

Example 4.15. The *Unclear Test* problem defines a set of test results, $Y=\{P,N\}$, and a set of health outcomes, $X=\{I,H\}$. We define $f^1\colon 1\to\mathbf{D}_\circ(X)$ by $f^1()=\frac{3}{4}[P]+\frac{1}{4}[N]$; we define $f^2\colon 1\to\mathbf{D}_\circ(Y)$ by $f^2()=\frac{1}{3}[I]+\frac{2}{3}[H]$; we define $f^3\colon Y\to\mathbf{D}_\circ(X)$ by $f^3(I)=\frac{3}{4}[P]+\frac{1}{4}[N]$ and $f^3(H)=\frac{1}{2}[P]+\frac{1}{2}[N]$; and, again, we define $f^4\colon X\times X\to\mathbf{D}_\circ(1)$ as an equality check: $f^4(x_1,x_2)=\top$ when $x_1=x_2$ and $f^4(x_1,x_2)=\bot$ when $x_1\neq x_2$. Rearranging inputs, we derive $f^2_{\sigma_2^\star}\colon X\to\mathbf{D}_\circ(X\times Y)$; we derive $f^3_{\sigma_3^\star}\colon X\times Y\to\mathbf{D}_\circ(X\times Y\times X)$; and we derive $f^4_{\sigma_4^\star}\colon X\times Y\times X\to\mathbf{D}_\circ(X\times Y\times X)$.

The two solutions to the *Unclear Test* problem (Section 1.2) correspond, at each line, to a factor of the following compositions — plus the final projection, $\pi\colon X\times Y\times X\to\mathbf{D}_\circ(X)$, which we omit.

$$\begin{aligned}(f^1\fatsemi f^2_{\sigma_2^\star}\fatsemi f^3_{\sigma_3^\star})\fatsemi f^4_{\sigma_4^\star}&=f^1\mathbin{;}\mathbf{D}_\circ f^2_{\sigma_2^\star}\mathbin{;}\mu\mathbin{;}\mathbf{D}_\circ f^3_{\sigma_3^\star}\mathbin{;}\mu\mathbin{;}\mathbf{D}_\circ f^4_{\sigma_4^\star}\mathbin{;}\mu\\&\neq f^1\mathbin{;}\mathbf{D}_\circ(f^2_{\sigma_2^\star}\mathbin{;}\mathbf{D}_\circ f^3_{\sigma_3^\star}\mathbin{;}\mu\mathbin{;}\mathbf{D}_\circ f^3_{\sigma_4^\star}\mathbin{;}\mu)\mathbin{;}\mu=f^1\fatsemi(f^2_{\sigma_2^\star}\fatsemi f^3_{\sigma_3^\star}\fatsemi f^4_{\sigma_4^\star})\end{aligned}$$

Remark 4.16. Magmoid expressions are not particularly pleasant to describe inference problems. We would rather describe problems explicitly in terms of variables, instead of compositions, identities, and finite functions. The next section adapts the monadic metalanguage [45, 70] to magmads.

4.4 The Magmadic metalanguage

Moggi's *system PL* [45] is a metalanguage for monadic computation, popularly known as *do-notation* [27]. Do-notation takes semantics over cartesian closed categories endowed with a strong monad. It consists of two rules, and it leaves variable management to the metatheory. The monadic metalanguage is implicitly right associative: variables are bound to the right. For magmads, this convention becomes a choice: we may want to associate to the left, and indeed we will see that this is sometimes desirable.

We start by recalling the definition of magmad in relative form (Definition 4.17) and introducing both left and right magmadic do-notation (Definition 4.19).

Definition 4.17 (Magmad in relative form)**.** A (Set-based) *magmad*, $(\mathbf{T}, \eta^{\mathbf{T}}, \beta^{\mathbf{T}})$, consists of an assignment on sets, $\mathbf{T}X$ for each set X, a natural family of functions $\eta^{\mathbf{T}} \colon X \to \mathbf{T}X$ and a natural family of functions $\beta^{\mathbf{T}} \colon \mathbf{T}X \times (X \to \mathbf{T}Y) \to \mathbf{T}Y$ (called *binding*), satisfying

(1) $\beta^{\mathbf{T}}(\eta^{\mathbf{T}}(x))(\lambda x.f(x)) = f(x)$, left unitality; and
(2) $\beta^{\mathbf{T}}(t)(\lambda x.\eta^{\mathbf{T}}(x)) = t$, right unitality.

A monad additionally satisfies associativity, $\beta^{\mathbf{T}}(t)(\lambda x.\beta^{\mathbf{T}}(f(x))(g)) = \beta^{\mathbf{T}}(\beta^{\mathbf{T}}(t)(\lambda x.f(x)))(\lambda y.g(y))$.

Example 4.18. The magmad of normalized distributions in relative form, $(\mathbf{D}_\circ, \eta^{\mathbf{D}_\circ}, \beta^{\mathbf{D}_\circ})$, is defined by $\eta^{\mathbf{D}_\circ}(x) = [x]$ and $\beta^{\mathbf{D}_\circ}(\sum_i \lambda_i\, [x_i])(x_i \mapsto \sum_j \rho_{ij}\, [y_j]) = \sum_{ij}(\lambda_i \rho_{ij} / \sum_{ij} \lambda_i \rho_{ij})\, [x_i y_j]$.

Definition 4.19 (Do notation)**.** The magmadic metalanguage is defined by a grammar containing *expressions*, *statements*, and *return*.

$$
\begin{aligned}
E &::= (\mathsf{rightDo}\ S_1 \dots S_n\ R) \mid (\mathsf{leftDo}\ \ S_1 \dots S_n\ R) \mid (f\ x_1 \dots x_n) \\
S &::= (x_1 \dots x_n) \leftarrow E \\
R &::= \mathsf{return}\ \ (x_1 \dots x_n)
\end{aligned}
$$

Right do notation inductively interprets *(i)* statements as bind; *(ii)* returns as unit.

$$
\begin{aligned}
(\mathsf{rightDo}\ \ (x_1 \dots x_n) \leftarrow e\ \ S_1 \dots S_k\ R) &\overset{(i)}{=} \beta^{\mathbf{T}}(e)\,(\lambda(x_1, \dots, x_n).(\mathsf{rightDo}\ S_1 \dots S_k\ R))\,; \\
(\mathsf{rightDo}\ \mathsf{return}\ \ (x_1 \dots x_n)) &\overset{(ii)}{=} \eta^{\mathbf{T}}(x_1, \dots, x_n).
\end{aligned}
$$

Left do notation inductively interprets (i, ii) the base cases of zero or one statements as right do notation, and (iii) any longer cases by accumulating the first two statements.

$$
\begin{aligned}
(\mathsf{leftDo}\ \ \mathsf{return}\ \ (x_1 \dots x_n)) &\overset{(i)}{=} \eta^{\mathbf{T}}(x_1, \dots, x_n); \\
(\mathsf{leftDo}\ \ (x_1 \dots x_n) \leftarrow e\ \mathsf{return}\ (y_1 \dots y_m)) &\overset{(ii)}{=} \beta^{\mathbf{T}}(e)(\lambda(x_1, \dots, x_n).\eta^{\mathbf{T}}(y_1, \dots, y_m)); \\
(\mathsf{leftDo}\ \ (x_1 \dots x_n) \leftarrow e_1 \quad & \\
\quad (y_1 \dots y_m) \leftarrow e_2 \quad & \overset{(iii)}{=} \quad (\mathsf{leftDo}\ \ (x_1 \dots x_n\ y_1 \dots y_m) \leftarrow (\mathsf{rightDo} \\
\quad S_1 \dots S_k\ R) \quad & \qquad (x_1 \dots x_n) \leftarrow e_1 \\
& \qquad (y_1 \dots y_m) \leftarrow e_2 \\
& \qquad \mathsf{return}\ \ (x_1 \dots x_n\ y_1 \dots y_m)) \\
& \quad S_1 \dots S_k\ R).
\end{aligned}
$$

Finally, function expressions are interpreted as function application, $(f\ x_1 \dots x_n) = f(x_1, ..., x_n)$.

Remark 4.20. By default, we use left do notation, $(\mathsf{do}\ S_1 \dots S_n\ R) = (\mathsf{leftDo}\ S_1 \dots S_n\ R)$.

Remark 4.21 (Implementing the magmadic metalanguage). Haskell's do-notation [27, 28] is right-associative by default: left-associativity is not part of the language. Imitating its syntax, we provide an implementation in Racket that relies on the macro facilities of the languages of the Lisp family.

4.5 Application — Modes of update

Updating a prior distribution with an observation can be done in different ways, depending on the structure of our model and the observation. Following Jacobs [31], we call them *modes of update*.

This section formalizes modes of update as non-associative phenomena. The magmadic metalanguage allows us to specify different modes of update: we compare Bayes' update to the filtering of impossible cases (Proposition 4.22). We compare Pearl's update to Jeffrey's update.

Three-statement case. Let us first clarify the anatomy of a basic Bayesian update. It contains three steps: (*i*) we have a prior, (*ii*) we apply a stochastic channel, (*iii*) and we observe some property through that channel: the association of these three steps yields two modes of update.

Proposition 4.22 (Bayes update)**.** *Let* prior$\colon 1 \to \mathbf{D}_\circ X$ *and* channel$\colon X \to \mathbf{D}_\circ Y$ *be total stochastic kernels; let* property$\colon Y \to \mathbf{D}_\circ 1$ *be a partial deterministic function. There are two distinct modes of update: Bayes' update and the removal of impossible cases (i.e., cases with probability zero).*

```
(leftDo  (x) ⟵ prior
         (y) ⟵ (channel x)
         () ⟵ (property y)
         return  (x))
```
$\neq$
```
(rightDo  (x) ⟵ prior
          (y) ⟵ (channel x)
          () ⟵ (property y)
          return  (x))
```

Proof. We have exactly 2 ways of associating three statements: 2 is the third Catalan number. These turn, by Definition 4.19, into the following formulas.

$$\begin{array}{l}\beta(\beta(\beta(\mathsf{prior})\\ \quad(\lambda x.\beta(\mathsf{channel}(x))(\lambda y.\eta(x,y)))\\ \quad\quad(\lambda x,y.\beta(\mathsf{property}(y))(\lambda.\eta(x,y))\\ \quad\quad\quad(\lambda x,y.\eta(x)))))\end{array} \neq \begin{array}{l}\beta(\mathsf{prior})(\lambda x.\\ \quad\beta(\mathsf{channel}(x))(\lambda y.\\ \quad\quad\beta(\mathsf{property}(y))(\lambda.\\ \quad\quad\quad\eta(x))))\end{array}$$

Let us write $\mathsf{prior} = \sum_i \lambda_i\,[x_i]$, where $\sum_i \lambda_i = 1$; let us write $\mathsf{channel}(x_i) = \sum_{j=0} \rho_{ij}\,[y_{ij}]$, where $\sum_j \rho_{ij} = 1$ and let $\mathsf{property}(y_{ij}) = u_{ij}$, where $u_{ij} = \bot$ or $u_{ij} = \top$. We compute explicitly the two possible updates by steps. The left-hand side can be computed as $\sum_i(\sum_j \lambda_i\rho_{ij}u_j/\sum_{i,j}\lambda_i\rho_{ij}u_j)\,[x_i]$.

$$\begin{array}{ll}\ell_1 = \mathsf{prior} & = \sum_i \lambda_i\,[x_i];\\ \ell_2 = \beta(\ell_1)(\lambda x.\mathsf{channel}(x)(\lambda y.\eta(x,y))) & = \sum_{i,j}\lambda_i\rho_{ij}\,[x_iy_j];\\ \ell_3 = \beta(\ell_2)(\lambda x,y.\mathsf{property}(y)(\lambda.\eta(x,y))) & = \sum_{i,j}(\lambda_i\rho_{ij}u_j/\sum_{i,j}\lambda_i\rho_{ij}u_j)\,[x_iy_j];\\ \ell_4 = \beta(\ell_3)(\lambda x,y.\eta(x)) & = \sum_i(\sum_j\lambda_i\rho_{ij}u_j/\sum_{i,j}\lambda_i\rho_{ij}u_j)\,[x_i];\end{array}$$

The right-hand side computes as $\sum_i \lambda_i \sum_j(\rho_{ij}u_j/\sum_j\rho_{ij}u_j)\,[x_i]$.

$$\begin{array}{ll}\kappa_1 = \eta(x_i) & = [x_i];\\ \kappa_2 = \beta(\mathsf{property}(y_j))(\lambda.\kappa_1) & = u_j\,[x_i];\\ \kappa_3 = \beta(\mathsf{channel}(x_i))(\lambda y.\kappa_2) & = \sum_j(\rho_{ij}u_j/\sum_j\rho_{ij}u_j)\,[x_i];\\ \kappa_4 = \beta(\mathsf{prior})(\lambda x.\kappa_3) & = \sum_i\lambda_i\sum_j(\rho_{ij}u_j/\sum_j\rho_{ij}u_j)\,[x_i];\end{array}$$

Note how the first corresponds to Bayes' update, while the second filters out impossible cases. We have $\ell_4 = \sum_i(\sum_{j,u_j=1}\lambda_i\rho_{ij}/\sum_{i,j,u_j=1}\lambda_i\rho_{ij})\,[x_i]$ but $\kappa_4 = \sum_{i,\sum_j\rho_{ij}u_j\neq 0}\lambda_i\,[x_i]$.[9] □

Example 4.23 (Monty Hall problem). The Monty Hall problem has two solutions, depending on associativity: they correspond to left do notation (leftDo/do) or right do notation (rightDo). The left-associative solution applies Bayes update and prefers to change doors ($\frac{1}{3}[M] + \frac{2}{3}[R]$); the right-associative solution filters the impossible case and is indifferent to changing doors ($\frac{1}{2}[M] + \frac{1}{2}[R]$).

```
do  (prize door) ⟵ (do
        (prize) ⟵ uniformDoor
        (door) ⟵ (host prize)
        return  (prize door))
    () ⟵ (observe L door)
    return  (prize)
```
$\neq$
```
do  (prize) ⟵ uniformDoor
    (door) ⟵ (do
        (door) ⟵ (host prize)
        () ⟵ (observe L door)
        return  (door))
    return  (prize)
```

[9] In both cases, we handle division by zero: whenever the denominator is zero, the numerator must have been zero.

Four-statement case. When the property depends on an extra stochastic parameter, there are instead four steps to consider: (*i*) we sample the parameter, (*ii*) we have a prior, (*iii*) we apply a stochastic channel, (*iv*) and we observe some property. The association of these four steps matters, and there are 5 ways of associating four statements. However, stochastic statements do associate, reducing the number of possibilities to four: of these, only two perform a Bayesian update with the three last statements.

Proposition 4.24 (Pearl's and Jeffrey's updates)**.** *Let* parameter$\colon 1 \to \mathbf{D}_\circ Z$*, and* prior$\colon 1 \to \mathbf{D}_\circ X$*, and* channel$\colon X \to \mathbf{D}_\circ Y$ *be stochastic and total functions let* property$\colon Y \times Z \to \mathbf{D}_\circ 1$ *be a partial and deterministic function. There are two modes of update extending Bayes' update: (i) Pearl's update and (ii) Jeffrey's update. Pearl's and Jeffrey's updates do not coincide in general.*

```
do (z) ⟵ parameter
   (x) ⟵ prior
   (y) ⟵ (channel x)
   () ⟵ (property y z)
   return (x)
```

≠

```
do (z) ⟵ parameter
   (x y) ⟵ (do (x) ⟵ prior
               (y) ⟵ (channel x)
               () ⟵ (property y z)
               return (x y))
   return (x)
```

Proof. One of our examples (the *Unclear Test problem* in Section 1.2) follows these two modes of update and obtains different results. We do not characterize these further. □

Remark 4.25 (Interpreting Pearl and Jeffrey). In the *Unclear Test problem* Section 1.2, Pearl's update uses a single normalization constant for both the positive-result and the negative-result branches; Jeffrey's update normalizes both separately. The greater relative likelihood of the positive branch makes Pearl's update yield a higher probability of illness. In the work of Jacobs [31], Pearl's update has been characterized as *increasing validity*, while Jeffrey's update as *decreasing divergence*.

5 Reasoning with the Metalanguage

This section introduces principles to reason about programs involving observations. Each axiom is based on one of the semantic principles we have established until now, except for the disintegration axiom, which we introduce.

Figure 3 lists the six axioms we will use: ① commutativity of observe and the Frobenius axiom, with an extra equation defining the notion of full support (justified in Section 5.1); ② the exchange axiom (justified in Section 5.2); ③ discarding of stochastic morphisms, copying of partial morphisms, and the normal magmad equation for arbitrary morphisms (justified in Section 5.3); ④ the disintegration axiom and its algebraic version (justified in Section 5.4); and ⑤ associativity, whenever F is partial, G is partial, or H is stochastic (justified in Proposition 4.12).

From these, we highlight how *commutativity* and *monoidality* — unlike in the monadic case — do not necessarily coincide for magmads: we introduce a notion of left-exchanging magmad (Definition 5.6).

5.1 Axiom — Partial Frobenius

We ask for the existence of a partial equality check, observe$\colon X \times X \to \mathbf{N}1$, which forces its two arguments to coincide: for the normalized distribution magmad, this is defined by observe$(x, y) = 0$ when $x \neq y$ and by observe$(x, x) = 1$ [] otherwise. Apart from commutativity, the only axiom that we ask of this morphism is that it locally forces the equality of its two arguments.

Definition 5.1. A magmad, $(\mathbf{N}, \mu^{\mathrm{N}}, \eta^{\mathrm{N}})$, is *partial Frobenius* whenever there exists, for each object X, a morphism, observe$\colon X \times X \to \mathbf{N}1$, satisfying *(i)* commutativity and *(ii)* the Frobenius axiom.

$$\boxed{(\text{observe } x\ y)} \overset{(i)}{=} \boxed{(\text{observe } y\ x)}$$

$$\boxed{\begin{array}{l}(\text{do } S_1 \ldots S_k \\ \quad () \longleftarrow (\text{observe } a\ b) \\ \quad (x_1 \ldots x_n) \longleftarrow (E\ a) \\ \quad S_{k+1} \ldots S_n\ R)\end{array}} \overset{(ii)}{=} \boxed{\begin{array}{l}(\text{do } S_1 \ldots S_k \\ \quad () \longleftarrow (\text{observe } a\ b) \\ \quad (x_1 \ldots x_n) \longleftarrow (E\ b) \\ \quad S_{k+1} \ldots S_n\ R)\end{array}}$$

Remark 5.2. Observe operators correspond to *partial Frobenius structures* in the sense of restriction categories [12] and *partial Markov categories* [10, 11]. The *comparator* is obtained from the observe morphisms and variable-copying. We use it to derive a notion of full support.

Definition 5.3 (Full support)**.** A Kleisli morphism, $E_{\text{full}}\colon X \to \mathbf{N}Y$, has *full support* whenever the following equation holds.

$$\boxed{\begin{array}{l}(\text{do } (y_1 \ldots y_m) \longleftarrow E_{full} \\ \quad () \longleftarrow (\text{observe } (x_1 \ldots x_n)\ (y_1 \ldots y_m)) \\ \quad \text{return } (x_1 \ldots x_n))\end{array}} = \boxed{(\text{do return } (x_1 \ldots x_n))}$$

5.2 Commutative and Left-Exchanging magmads

Programming with commutative magmads raises subtleties. For instance, we would expect that, for probabilistic programs, exchanging the lines of a program must not change its meaning, as long as the lines do not depend on each other [34, 60, 71].[10] That is, whenever $x_1, \ldots, x_n$ are free in E_2 and $y_1, \ldots, y_m$ are free in E_1, the following *exchange rule* holds.

$$\boxed{\begin{array}{l}(\text{do } S_1 \ldots S_k \\ \quad (x_1 \ldots x_n) \longleftarrow E_1 \\ \quad (y_1 \ldots y_m) \longleftarrow E_2 \\ \quad S_{k+1} \ldots S_n\ R)\end{array}} = \boxed{\begin{array}{l}(\text{do } S_1 \ldots S_k \\ \quad (y_1 \ldots y_m) \longleftarrow E_2 \\ \quad (x_1 \ldots x_n) \longleftarrow E_1 \\ \quad S_{k+1} \ldots S_n\ R)\end{array}}$$

As it happens for monads, every Set-magmad is *strong*, meaning that there exist canonical maps $\psi_\ell\colon X \otimes \mathbf{T}(Y) \to \mathbf{T}(X \otimes Y)$ and $\psi_r\colon \mathbf{T}(X) \otimes Y \to \mathbf{T}(X \otimes Y)$. In some cases, the two ways in which they induce a map $\psi\colon \mathbf{T}(X) \otimes \mathbf{T}(Y) \to \mathbf{T}(X \otimes Y)$ coincide: there exists a canonical way of interpreting two effects in parallel; we say that the magmad is commutative (Definition 5.4).

Perhaps surprisingly, the *exchange rule* in left-do notation does not follow from commutativity but is a stronger property of its own: left-exchange (Definition 5.6). Commutativity only implies left-exchange in the presence of associativity (Proposition 5.7); for magmads, left-exchange becomes an independent axiom. We prove that sesquilaws induce left-exchanging magmads (Proposition 5.8).

Definition 5.4 (Commutative magmad)**.** A *commutative magmad* is a magmad, $(\mathbf{T}, \mu^{\mathrm{T}}, \eta^{\mathrm{T}})$, such that the following diagram commutes.

$$\begin{array}{ccccc} \mathbf{T}X \otimes \mathbf{T}Y & \xrightarrow{\mathbf{T}\psi_\ell} & \mathbf{T}(\mathbf{T}X \otimes Y) & \xrightarrow{\mathbf{T}\psi_r} & \mathbf{TT}(X \otimes Y) \\ \downarrow{\scriptstyle \mathbf{T}\psi_r} & & & & \downarrow{\scriptstyle \mu^{\mathrm{T}}} \\ \mathbf{T}(X \otimes \mathbf{T}Y) & \xrightarrow{\mathbf{T}\psi_\ell} & \mathbf{TT}(X \otimes Y) & \xrightarrow{\mu^{\mathrm{T}}} & \mathbf{T}(X \otimes Y) \end{array}$$

[10] This property is usually called *commutativity* itself.

Remark 5.5 (Monoidal magmad). The general definition of monoidal magmad — the generalization outside the category of sets — is not required for our development. All of the magmads and monads of this text are monoidal in this general sense.

Definition 5.6 (Left-exchanging magmad)**.** A *left-exchanging magmad* is a commutative magmad, $(\mathbf{T}, \mu^{\mathbf{T}}, \eta^{\mathbf{T}})$, such that, moreover, the following two diagrams commute.

$$\begin{array}{ccccc}
\mathbf{T}(\mathbf{T}X \otimes \mathbf{T}Y) & \xrightarrow{\mathbf{T}\psi} & \mathbf{T}\mathbf{T}(X \otimes Y) & \xrightarrow{\mu^{\mathbf{T}}} & \mathbf{T}(X \otimes Y) \\
{\scriptstyle \mathbf{T}\psi_\ell}\downarrow & ① & & & \uparrow{\scriptstyle \mu^{\mathbf{T}}} \\
\mathbf{T}\mathbf{T}(X \otimes \mathbf{T}Y) & \underset{\mu^{\mathbf{T}}}{\rightrightarrows} & \mathbf{T}(X \otimes \mathbf{T}Y) & \underset{\mathbf{T}\psi_r}{\rightrightarrows} & \mathbf{T}\mathbf{T}(X \otimes Y)
\end{array}
\qquad
\begin{array}{ccccc}
\mathbf{T}(\mathbf{T}X \otimes \mathbf{T}Y) & \xrightarrow{\mathbf{T}\psi} & \mathbf{T}\mathbf{T}(X \otimes Y) & \xrightarrow{\mu^{\mathbf{T}}} & \mathbf{T}(X \otimes Y) \\
{\scriptstyle \mathbf{T}\psi_r}\downarrow & ② & & & \uparrow{\scriptstyle \mu^{\mathbf{T}}} \\
\mathbf{T}\mathbf{T}(\mathbf{T}X \otimes Y) & \underset{\mu^{\mathbf{T}}}{\rightrightarrows} & \mathbf{T}(\mathbf{T}X \otimes Y) & \underset{\mathbf{T}\psi_\ell}{\rightrightarrows} & \mathbf{T}\mathbf{T}(X \otimes Y)
\end{array}$$

Proposition 5.7. *A monad is commutative if and only if it is left-exchanging. Not every commutative magmad is a left-exchanging magmad.*

Proof sketch. For the monad case, we employ naturality, the associativity of the monad, and the definition of the joint strength to prove that commutative implies left-exchanging.

To prove that associativity is required in the magmadic case, consider the *average-writer magmad*, $(\mathbf{Q}, \mu^{\mathbf{Q}}, \eta^{\mathbf{Q}})$, given by the functor $\mathbf{Q}X = X \times \mathbf{M}\mathbb{Q}^+$, with the multiplication defined by $\mu^{\mathbf{Q}}(x, q_1, q_2) = (x, (q_1 + q_2)/2)$ — with $\mu^{\mathbf{Q}}(x, \bot, q_2) = (x, q_2)$ and $\mu^{\mathbf{Q}}(x, q_1, \bot) = (x, q_1)$ — and the unit defined by $\eta^{\mathbf{Q}}(x) = (x, \bot)$. This is a commutative magmad but not a left-exchanging magmad. □

Proposition 5.8. *Sesquilaws induce left-exchanging magmads.*

5.3 Affine, Relevant, and Normal Magmads

The category of sets and functions has the structure to copy and discard information; explicitly, there exist transformations $\delta\colon X \to X \times X$ and $\varepsilon\colon X \to 1$ that are natural. Monads and magmads do not necessarily preserve this structure: some effects cannot be copied or discarded without changing their meaning; the following two equations (①, ②) do not always hold.

For instance, stochasticity cannot be copied (② fails) — throwing two coins is different from throwing a single coin and copying the result; and partiality cannot be discarded (① fails) — even if we discard the result of a partial function, its domain remains unaltered. Instead, stochasticity can be discarded ①, and partiality can be copied ② (Proposition 5.10): monads that can be discarded are called *affine*; monads that can be copied are called *relevant* (Definition 5.9).

Definition 5.9 (Affine & relevant magmads, cf. [29])**.** An *affine magmad* and a *relevant magmad* are commutative magmads, $(\mathbf{T}, \mu^{\mathbf{T}}, \eta^{\mathbf{T}}, \psi^{\mathbf{T}})$, such that — respectively — the first or the second of the following diagrams commute.

$$\begin{array}{ccc}
\mathbf{T}X & \xrightarrow{\mathbf{T}\varepsilon} & \mathbf{T}1 \\
{\scriptstyle \varepsilon}\downarrow & ① & \nearrow{\scriptstyle \eta^{\mathbf{T}}} \\
1 & &
\end{array}
\qquad
\begin{array}{ccc}
\mathbf{T}X & \xrightarrow{\mathbf{T}\delta} & \mathbf{T}(X \otimes X) \\
{\scriptstyle \delta}\downarrow & ② & \nearrow{\scriptstyle \psi} \\
\mathbf{T}X \otimes \mathbf{T}X & &
\end{array}$$

Equivalently, the first or — respectively — the second of the following equations hold for any $F \in \mathbf{T}X$.

```
(do return ())
```
$\overset{①}{=}$
```
(do (x₁ ... xₙ) ⟵ F
    return ())
```

```
(do (x₁ ... xₙ) ⟵ F
    (y₁ ... yₘ) ⟵ F
    return (x₁ ... xₙ y₁ ... yₘ))
```
$\overset{②}{=}$
```
(do (x₁ ... xₙ) ⟵ F
    return (x₁ ... xₙ x₁ ... xₙ))
```

Proposition 5.10. *The distribution monad* $(\mathbf{D}, \mu^{\mathrm{D}}, \eta^{\mathrm{D}})$ *is affine; the maybe monad* $(\mathbf{M}, \mu^{\mathrm{M}}, \eta^{\mathrm{M}})$ *is relevant. The magmad of normalized distributions,* $(\mathbf{MD}, \mu_{\circ}^{\mathrm{MD}}, \eta_{\circ}^{\mathrm{MD}})$ *is a commutative magmad that is not affine nor relevant.*

Normalized distributions form a magmad that is not affine nor relevant. Totality can be refined to account for updates: the resulting notion is that of a normal magmad — or, in the terminology of *partial Markov categories*, a magmad whose effects are quasitotal [10, 11, 54]. This equation is missing in any non-normalized probabilistic semantics (Corollary 5.13).

Definition 5.11 (Normal magmad)**.** A *normal magmad* is a commutative magmad, $(\mathbf{N}, \mu^{\mathbf{N}}, \eta^{\mathbf{N}}, \psi^{\mathbf{N}})$, such that the following diagram commutes.

$$\begin{array}{ccc} \mathbf{N}X & \longrightarrow & \mathbf{N}X \\ {\scriptstyle\delta}\downarrow & \textcircled{3} & \uparrow{\scriptstyle\mathbf{N}(X\otimes\varepsilon)} \\ \mathbf{N}X \otimes \mathbf{N}Y & \xrightarrow{\psi} & \mathbf{N}(X \otimes X) \end{array}$$

Equivalently, the following equation holds for any expression $F \in \mathbf{N}(X_1 \times \ldots \times X_n)$.

```
(do (x_1 ... x_n) ⟵ F
    (y_1 ... y_m) ⟵ F
    return (x_1 ... x_n))
```

$\overset{(3)}{=}$

```
(do (x_1 ... x_n) ⟵ F
    return (x_1 ... x_n))
```

Proposition 5.12 (Affine-relevant sesquilaw)**.** *An* affine-relevant sesquilaw *is a monoidal sesquilaw,* $(\mathbf{S}, \mathbf{T}, m, n)$*, of an affine monad,* $(\mathbf{T}, \mu^{\mathbf{T}}, \eta^{\mathbf{T}})$ *over a relevant monad,* $(\mathbf{S}, \mu^{\mathbf{S}}, \eta^{\mathbf{S}})$*. Any affine-relevant sesquilaw* $(\mathbf{S}, \mathbf{T}, m, n)$ *induces a quasitotal magmad* $(\mathbf{ST}, \mu_n^{\mathbf{ST}}, \eta_n^{\mathbf{ST}})$.

Corollary 5.13. *The magmad of normalized distributions,* $(\mathbf{MD}, \mu_{\circ}^{\mathrm{MD}}, \eta_{\circ}^{\mathrm{MD}})$ *is a normal magmad. The monad of subdistributions,* $(\mathbf{DM}, \mu_{\bullet}^{\mathrm{MD}}, \eta_{\bullet}^{\mathrm{MD}})$ *is not a normal monad.*

5.4 Axiom — Disintegration

Disintegration was introduced by Cho and Jacobs [6] as a synthetic axiom for extracting a conditional channel, $d(f) \colon X \to \mathbf{N}Y$, from a joint distribution, $f \in \mathbf{N}(X \times Y)$ (Definition 5.14). For the normalized distribution magmad, it is witnessed by the existence of conditional distributions.

A limitation of its formulation is that it is a non-equational axiom: it postulates that there exists such disintegration, but not how to construct it; in fact, it is unique only in the degenerate case where the all projections coincide [17, Proposition 11.15]. However, in some cases, a disintegration can be constructed, and we may substitute the existential axiom for an algebraic one (Definition 5.15, cf. [40, Proposition 109]).[11]

Definition 5.14 (Disintegration, (cf. [6, Definition 3.5]))**.** A magmad, $(\mathbf{N}, \mu^{\mathbf{N}}, \eta^{\mathbf{N}})$ satisfies the *disintegration axiom* whenever, for every expression $E \in \mathbf{N}(X \times Y)$, there exists some function $F \colon X \to \mathbf{N}(Y)$ such that the following diagram commutes.

```
(do (x_1 ... x_n y_1 ... y_m) ⟵ E
    return (x_1 ... x_n y_1 ... y_m))
```

$\overset{(i)}{=}$

```
(do (x_1 ... x_n y_1 ... y_m) ⟵ E
    (y_1 ... y_m) ⟵ (F x_1 ... x_n)
    return (x_1 ... x_n y_1 ... y_m))
```

In that particular case, we say that F is a disintegration of E.

[11] Both the existential and the algebraic formulations of disintegration can be generalized to monoidal magmoids. For the purposes of this text, we work exclusively with Kleisli magmoids over the category sets: instead of dealing with arbitrary channels $A \to \mathbf{N}(X \times Y)$, we simplify the definition using that a Kleisli map is determined by its value on each element.

Definition 5.15 (Algebraic disintegration)**.** A magmad, $(\mathbf{N}, \mu^{\mathbf{N}}, \eta^{\mathbf{N}})$ satisfies the *algebraic disintegration* axiom whenever, for each expression $E \in \mathbf{N}(X \times Y)$, the composite $\partial E \colon X \to \mathbf{N}Y$, defined as follows, satisfies the previous equation in Definition 5.14.*(i)*.

$$\boxed{(\partial E\ x_1 \ldots x_n)} \overset{(ii)}{=} \boxed{\begin{array}{l}(\text{do } (a_1 \ldots a_n\ y_1 \ldots y_m) \leftarrow E\\ \quad () \leftarrow (\text{observe } (x_1 \ldots x_n)\ (a_1 \ldots a_n))\\ \quad \text{return } (y_1 \ldots y_m))\end{array}}$$

5.5 Reasoning — Pearl's Front-door and Back-door criteria

The axioms of the probabilistic magmadic metalanguage (Figure 3) allow us to prove basic lemmas about synthetic causal inference (cf. [51]): the language naturally distinguishes between observations and interventions as different parenthesizations.

Remark 5.16. Jacobs, Széles, and Stein [33] have recently shown how partial Markov categories [10], extended with normalization boxes [40], can be applied to synthetic causality (cf. [52, 72]). We prove similar results but, for the first time, satisfying normality (Corollary 5.13) and in a metalanguage with axioms: this seems to render the proofs slightly simpler and paves the way for generalization.

Theorem 5.17 (Front-door adjustment formula [51])**.** *Given a magmad,* $(\mathbf{N}, \mu^{\mathbf{N}}, \eta^{\mathbf{N}})$*, satisfying the principles in Figure 3, let an expression,* $P \in \mathbf{N}(X \times Y \times Z)$*, admit the following factorization into total morphisms —* $G \colon 1 \to \mathbf{N}U$*,* $S \colon U \to \mathbf{N}X$*,* $T \colon X \to \mathbf{N}Y$*, and* $C \colon Y \times U \to \mathbf{N}Z$ *— where, moreover, both* T *and* $(G \fatsemi S)$ *have full support.*

$$\boxed{\begin{array}{l}(\text{do } (s\ t\ c) \leftarrow P\\ \quad \text{return } (s\ t\ c))\end{array}} = \boxed{\begin{array}{l}(\text{do } (g) \leftarrow G\\ \quad (s) \leftarrow (S\ g)\\ \quad (t) \leftarrow (T\ s)\\ \quad (c) \leftarrow (C\ t\ g)\\ \quad \text{return } (s\ t\ c))\end{array}}$$

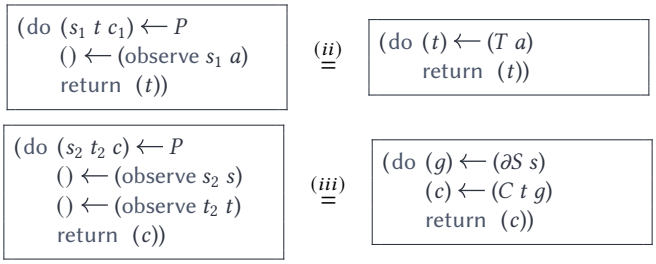
Then, the following equation holds. We define (frontDoor $P\ a$) *as its right-hand side.*

$$\boxed{\begin{array}{l}(\text{do } (g) \leftarrow G\\ \quad (t) \leftarrow (T\ a)\\ \quad (c) \leftarrow (C\ t\ g)\\ \quad \text{return } (c))\end{array}} = \boxed{\begin{array}{l}(\text{do } (s\ t_0\ c_0) \leftarrow P\\ \quad (t) \leftarrow (\text{do } (s_1\ t\ c_1) \leftarrow P\\ \qquad\qquad () \leftarrow (\text{observe } s_1\ a)\\ \qquad\qquad \text{return } (t))\\ \quad (c) \leftarrow (\text{do } (s_2\ t_2\ c) \leftarrow P\\ \qquad\qquad () \leftarrow (\text{observe } s_2\ s)\\ \qquad\qquad () \leftarrow (\text{observe } t_2\ t)\\ \qquad\qquad \text{return } (c))\\ \quad \text{return } (c))\end{array}}$$

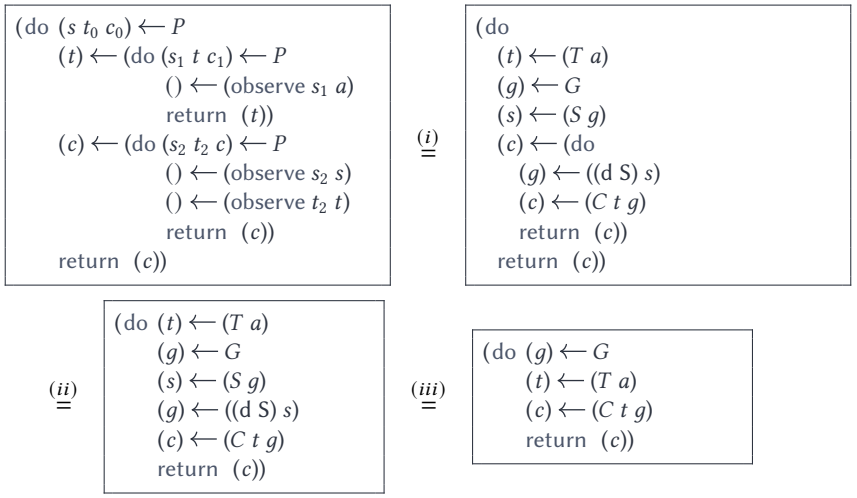
In other words, an intervention setting s *to* a *can be derived as a function of the observational data.*

Proof sketch. We simplify each one of the three parts of the equation. For the first part, *(i)* we use the hypothesis, and totality of C. For the second part, *(ii)* we also use Frobenius and the full-support assumption. For the third part, *(iii)* we employ the disintegration axiom. See the extended version for details [37].

$$\boxed{\begin{array}{l}(\text{do } (s\ t_0\ c_0) \leftarrow P\\ \quad \text{return } (s))\end{array}} \overset{(i)}{=} \boxed{\begin{array}{l}(\text{do } (g) \leftarrow G\\ \quad (s) \leftarrow (S\ g)\\ \quad \text{return } (s))\end{array}}$$

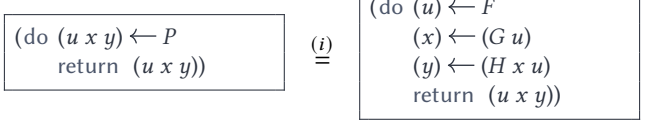


With these equations, let us address the main claim. We *(i)* substitute according to the previous three equations, *(ii)* use the totality of g and c, and *(iii)* apply the disintegration axiom.

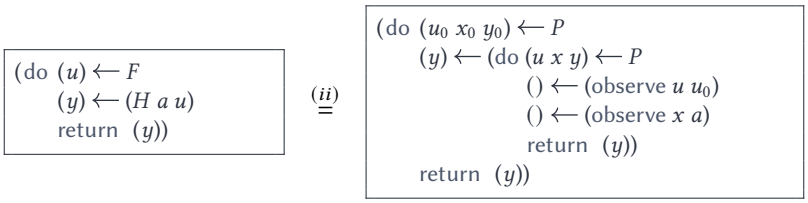


□

Theorem 5.18 (Back-door adjustment formula [51]). *Given a magmad,* $(\mathbf{N}, \mu^{\mathbf{N}}, \eta^{\mathbf{N}})$*, satisfying the principles in Figure 3, let an expression,* $P \in \mathbf{N}(U \times X \times Y)$*, admit the following factorization into total morphisms — where, moreover,* F *has full support.*



Then, the following equation holds. We define (backDoor P a) *as the right-hand side.*



Proof sketch. We apply the same reasoning principles as for the front-door formula. See the extended version for details [37]. □

From this perspective, causal inference is the resolution of equations in the magmoidal metalanguage of the normalized distribution monad. Alternatively to most probabilistic programming literature, we do not use a normalization operator, but an ability to modulate associativity.

6 Conclusions

Normalization has been under-explored in denotational probabilistic semantics; most semantics either ignore normalization — preferring substochastic kernels or unnormalized kernels instead — or treat it as an ad-hoc semantic feature. We have introduced the magmad of normalized distributions as a first example of *normalized-by-construction* denotational semantics.

Our proposed explanation for this gap is that normalization arises from a failure of associativity, and failures of associativity are counterintuitive. For instance, to solve an inference problem, one must (1) set up a prior distribution, (2) compute a stochastic kernel, and (3) update the prior with the observation; but the result depends on how we associate these, on *when to normalize*. Normalized probablistic semantics needs to clearly address this point, and a revision of existing probabilitic semantics to understand their potential non-associative behaviour is warranted.

Non-associativity has conceptual advantages: it allows us to derive multiple properties of normalization from a few axioms and it allows us to distinguish between observations and interventions by reparenthesizing. That normalization is not a distributive law may be folklore, but a systematic study was missing. We went one step further and classified this particular failure of normalization by introducing sesquilaws: quasidistributive laws that are multiplicative only up to idempotent.

Sesquilaws that are monoidal induce not only monoidal magmads but left-exchanging magmads, a novel notion refining commutativity the non-associative case. For left-exchanging magmads, we provided a variant of the monadic metalanguage that is not associative but preserves the line-exchange property. Remarkably, these seem to mostly coincide — although not exactly — with the existing string diagrammatic calculi for normalization over *partial Markov categories* [10, 33, 40].

Followin Jacobs' description of Pearl's and Jeffrey's modes of update [31], we have claimed that the difference between modes of update is a non-associative phenomenon. Finally, we proposed a different approach to categorical causality where solving causality problems corresponds to solving equations on a magmadic metalanguage. Further work on automating these solutions — e.g. via rewriting, showing completeness for *identifiability* — is warranted and not covered by this manuscript; it is a promising avenue for a categorical semantics of causal programming.

More generally, we reassert — as weak distributive laws [3, 22, 64], duploids [41, 46], or mass-chance interpretations do [66] — that failing distributive laws contain, in many cases, mathematical structure worth studying on its own.

Data availability statement

This article is accompanied by a code artifact in the Racket programming language implementing the magmadic metalanguage, exemplifying it with the Monty Hall problem and the Unclear Test problem, and providing an example of the magmadic front-door adjustment formula. The artifact can be found at the following permanent link.

- https://zenodo.org/records/23336667?preview_file=README.md

Disclosure of Use of Artificial Intelligence in Research

The authors declare no use of generative artificial intelligence to conduct research, to generate, draft or rewrite the contents of the article, nor to generate code for the accompanying code artifact.

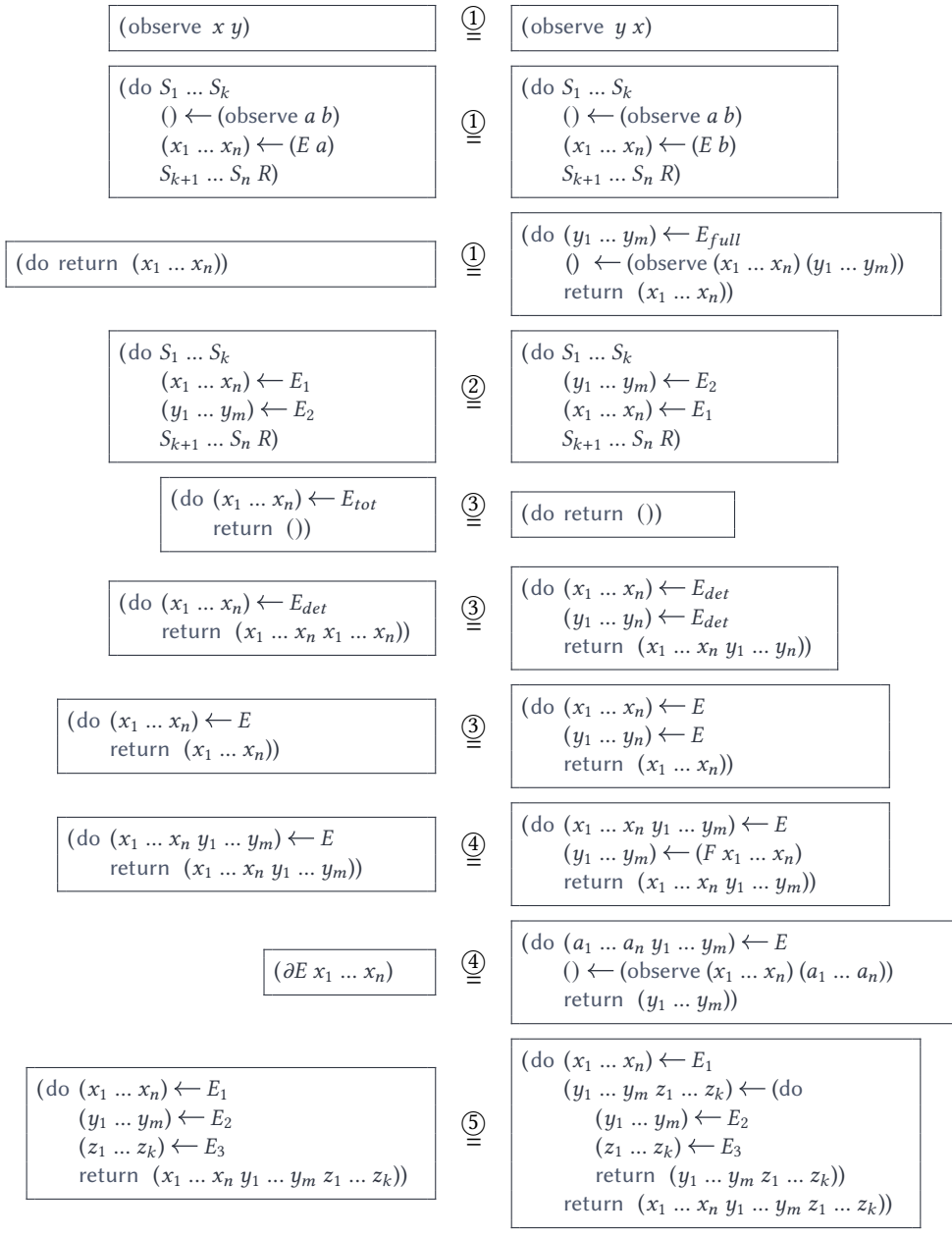


Fig. 3. Principles of the normalized stochastic magmadic metalanguage. ① Properties of observe, valid when E_{full} has full support, justified in Section 5.1. ② Exchange, justified in Section 5.2. ③ Valid when E_{det}, E_{tot} and E are total, deterministic and arbitrary, respectively, as justified in Section 5.3. ④ Definition of disintegration, justified in Section 5.4. ⑤ Valid when either E_1 is a partial function, E_2 is a partial function, or E_3 is a stochastic function, as justified in Proposition 4.12.

References

[1] Michael Barr and Charles Wells. 2000. *Toposes, triples, and theories.* Springer-Verlag.

[2] Jon Beck. 1969. Distributive laws. In *Seminar on triples and categorical homology theory.* Springer, 119–140.

[3] Gabriella Böhm. 2010. The weak theory of monads. *Advances in Mathematics* 225, 1 (2010), 1–32.

[4] Filippo Bonchi, Ana Sokolova, and Valeria Vignudelli. 2022. The Theory of Traces for Systems with Nondeterminism, Probability, and Termination. *Logical Methods in Computer Science* 18, 2 (2022). doi:10.46298/LMCS-18(2:21)2022

[5] Johannes Borgström, Ugo Dal Lago, Andrew D. Gordon, and Marcin Szymczak. 2016. A lambda-calculus foundation for universal probabilistic programming. In *Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming, ICFP 2016, Nara, Japan, September 18-22, 2016*, Jacques Garrigue, Gabriele Keller, and Eijiro Sumii (Eds.). ACM, 33–46. doi:10.1145/2951913.2951942

[6] Kenta Cho and Bart Jacobs. 2019. Disintegration and Bayesian Inversion via String Diagrams. *Mathematical Structures in Computer Science* (March 2019), 1–34. arXiv:1709.00322 doi:10.1017/S0960129518000488

[7] Corina Cîrstea, Lawrence S. Moss, Victoria Noquez, Todd Schmid, Alexandra Silva, and Ana Sokolova. 2025. A Complete Inference System for Probabilistic Infinite Trace Equivalence. In *EACSL Annual Conference on Computer Science Logic (CSL)*. doi:10.4230/LIPICS.CSL.2025.30

[8] Swaraj Dash, Younesse Kaddar, Hugo Paquet, and Sam Staton. 2023. Affine Monads and Lazy Structures for Bayesian Programming. *Proc. ACM Program. Lang.* 7, POPL (2023), 1338–1368. doi:10.1145/3571239

[9] Yuxin Deng, Rob J. van Glabbeek, Matthew Hennessy, and Carroll Morgan. 2009. Testing Finitary Probabilistic Processes. In *CONCUR 2009 - Concurrency Theory, 20th International Conference, CONCUR 2009, Bologna, Italy, September 1-4, 2009. Proceedings (Lecture Notes in Computer Science, Vol. 5710)*, Mario Bravetti and Gianluigi Zavattaro (Eds.). Springer, 274–288. doi:10.1007/978-3-642-04081-8_19

[10] Elena Di Lavore and Mario Román. 2023. Evidential Decision Theory via Partial Markov Categories. In *38th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2023, Boston, MA, USA, June 26-29, 2023*. IEEE, 1–14. doi:10.1109/LICS56636.2023.10175776

[11] Elena Di Lavore, Mario Román, and Paweł Sobociński. 2025. Partial Markov categories. *arXiv preprint* (2025). doi:10.48550/arXiv.2502.03477

[12] Ivan Di Liberti, Fosco Loregiàn, Chad Nester, and Pawel Sobocinski. 2021. Functorial semantics for partial theories. *Proc. ACM Program. Lang.* 5, POPL (2021), 1–28. doi:10.1145/3434338

[13] Thomas Ehrhard, Michele Pagani, and Christine Tasson. 2011. The Computational Meaning of Probabilistic Coherence Spaces. In *Proceedings of the 26th Annual IEEE Symposium on Logic in Computer Science, LICS 2011, June 21-24, 2011, Toronto, Ontario, Canada*. IEEE Computer Society, 87–96. doi:10.1109/LICS.2011.29

[14] Thomas Ehrhard, Michele Pagani, and Christine Tasson. 2017. Measurable cones and stable, measurable functions: a model for probabilistic higher-order programming. *Proceedings of the ACM on Programming Languages* 2, POPL (2017), 1–28. doi:10.1145/3158147

[15] Claudia Faggian and Simona Ronchi Della Rocca. 2019. Lambda Calculus and Probabilistic Computation. In *34th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2019, Vancouver, BC, Canada, June 24-27, 2019*. IEEE, 1–13. doi:10.1109/LICS.2019.8785699

[16] Brendan Fong. 2013. Causal Theories: A Categorical Perspective on Bayesian Networks. *Master's Thesis, University of Oxford. ArXiv preprint arXiv:1301.6201* (2013).

[17] Tobias Fritz. 2020. A Synthetic Approach to Markov Kernels, Conditional Independence and Theorems on Sufficient Statistics. *Advances in Mathematics* 370 (2020), 107239. arXiv:1908.07021

[18] Tobias Fritz, Tomás Gonda, Antonio Lorenzin, Paolo Perrone, and Areeb Shah-Mohammed. 2025. Empirical Measures and Strong Laws of Large Numbers in Categorical Probability. *CoRR* abs/2503.21576 (2025). arXiv:2503.21576 doi:10.48550/ARXIV.2503.21576

[19] Tobias Fritz and Andreas Klingler. 2023. The d-Separation Criterion in Categorical Probability. *J. Mach. Learn. Res.* 24 (2023), 46:1–46:49. https://jmlr.org/papers/v24/22-0916.html

[20] Tobias Fritz, Paolo Perrone, and Sharwin Rezagholi. 2021. Probability, valuations, hyperspace: Three monads on Top and the support as a morphism. *Mathematical Structures in Computer Science* 31, 8 (2021), 850–897.

[21] Richard Garner. 2018. Abstract hypernormalisation, and normalisation-by-trace-evaluation for generative systems. *arXiv preprint arXiv:1811.02710* (2018). arXiv:1811.02710

[22] Richard Garner. 2020. The Vietoris monad and weak distributive laws. *Applied Categorical Structures* 28, 2 (2020), 339–354.

[23] Michèle Giry. 1982. A categorical approach to probability theory. In *Categorical aspects of topology and analysis.* Springer, 68–85.

[24] Petr Viktorovich Golubtsov. 1999. Axiomatic description of categories of information transformers. *Problemy Peredachi Informatsii* 35, 3 (1999), 80–98.

[25] Noah D. Goodman, Vikash Mansinghka, Daniel M. Roy, Kallista A. Bonawitz, and Joshua B. Tenenbaum. 2012. Church: a language for generative models. *CoRR* abs/1206.3255 (2012). arXiv:1206.3255 http://arxiv.org/abs/1206.3255

[26] Alexandre Goy and Daniela Petrisan. 2020. Combining Weak Distributive Laws: Application to Up-To Techniques. *CoRR* abs/2010.00811 (2020). arXiv:2010.00811

[27] Paul Hudak, John Hughes, Simon L. Peyton Jones, and Philip Wadler. 2007. A history of Haskell: being lazy with class. In *Proceedings of the Third ACM SIGPLAN History of Programming Languages Conference (HOPL-III), San Diego, California, USA, 9-10 June 2007*, Barbara G. Ryder and Brent Hailpern (Eds.). ACM, 1–55. doi:10.1145/1238844.1238856

[28] Paul Hudak, Simon L. Peyton Jones, Philip Wadler, Brian Boutel, Jon Fairbairn, Joseph H. Fasel, María M. Guzmán, Kevin Hammond, John Hughes, Thomas Johnsson, Richard B. Kieburtz, Rishiyur S. Nikhil, Will Partain, and John Peterson. 1992. Report on the Programming Language Haskell, A Non-strict, Purely Functional Language. *ACM SIGPLAN Notices* 27, 5 (1992), 1. doi:10.1145/130697.130699

[29] Bart Jacobs. 1994. Semantics of weakening and contraction. *Annals of Pure and Applied Logic* 69, 1 (1994), 73–106. doi:10.1016/0168-0072(94)90020-5

[30] Bart Jacobs. 2017. Hyper Normalisation and Conditioning for Discrete Probability Distributions. *Log. Methods Comput. Sci.* 13, 3 (2017). doi:10.23638/LMCS-13(3:17)2017

[31] Bart Jacobs. 2019. The Mathematics of Changing One's Mind, via Jeffrey's or via Pearl's Update Rule. *J. Artif. Intell. Res.* 65 (2019), 783–806. doi:10.1613/jair.1.11349

[32] Bart Jacobs, Aleks Kissinger, and Fabio Zanasi. 2021. Causal inference via string diagram surgery: A diagrammatic approach to interventions and counterfactuals. *Mathematical Structures in Computer Science* 31, 5 (2021), 553–574.

[33] Bart Jacobs, Márk Széles, and Dario Stein. 2025. Compositional inference for Bayesian networks and causality, In Mathematical Foundations of Programming Semantics. *Proceedings of MFPS*.

[34] Xiaodong Jia, Bert Lindenhovius, Michael W. Mislove, and Vladimir Zamdzhiev. 2021. Commutative Monads for Probabilistic Programming Languages. In *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021*. IEEE, 1–14. doi:10.1109/LICS52264.2021.9470611

[35] Dexter Kozen. 1981. Semantics of Probabilistic Programs. *J. Comput. Syst. Sci.* 22, 3 (1981), 328–350. doi:10.1016/0022-0000(81)90036-2

[36] Ugo Dal Lago and Margherita Zorzi. 2012. Probabilistic operational semantics for the lambda calculus. *RAIRO Theor. Informatics Appl.* 46, 3 (2012), 413–450. doi:10.1051/ITA/2012012

[37] Elena Di Lavore, Mario Román, and Márk Széles. 2026. Normalized Probabilistic Semantics is Not Associative. arXiv:2510.01131 [math.CT] https://arxiv.org/abs/2510.01131

[38] F William Lawvere. 1962. The category of probabilistic mappings. *Unpublished preprint* (1962). Preprint available online, at https://ncatlab.org/nlab/files/lawvereprobability1962.pdf.

[39] Jack Liell-Cock and Sam Staton. 2024. Compositional imprecise probability. *CoRR* abs/2405.09391 (2024). arXiv:2405.09391 doi:10.48550/ARXIV.2405.09391

[40] Robin Lorenz and Sean Tull. 2023. Causal models in string diagrams. arXiv:2304.07638

[41] Éléonore Mangel, Paul-André Melliès, and Guillaume Munch-Maccagnoni. 2025. Classical notions of computation and the Hasegawa-Thielecke theorem. *arXiv preprint arXiv:2502.13033* (2025).

[42] Dylan McDermott and Tarmo Uustalu. 2022. What Makes a Strong Monad? *Electronic Proceedings in Theoretical Computer Science* 360 (2022), 113–133. doi:10.4204/eptcs.360.6

[43] Jan-Willem van de Meent, Brooks Paige, Hongseok Yang, and Frank Wood. 2018. An Introduction to Probabilistic Programming. *CoRR* abs/1809.10756 (2018). arXiv:1809.10756

[44] Matteo Mio, Ralph Sarkis, and Valeria Vignudelli. 2021. Combining Nondeterminism, Probability, and Termination: Equational and Metric Reasoning. In *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021*. IEEE, 1–14. doi:10.1109/LICS52264.2021.9470717

[45] Eugenio Moggi. 1991. Notions of Computation and Monads. *Inf. Comput.* 93, 1 (1991), 55–92. doi:10.1016/0890-5401(91)90052-4

[46] Guillaume Munch-Maccagnoni. 2013. *Syntax and Models of a non-Associative Composition of Programs and Proofs. (Syntaxe et modèles d'une composition non-associative des programmes et des preuves).* Ph. D. Dissertation. Paris Diderot University, France. https://tel.archives-ouvertes.fr/tel-00918642

[47] Praveen Narayanan, Jacques Carette, Wren Romano, Chung-chieh Shan, and Robert Zinkov. 2016. Probabilistic inference by program transformation in Hakaru (system description). In *International Symposium on Functional and Logic Programming - 13th International Symposium, FLOPS 2016, Kochi, Japan, March 4-6, 2016, Proceedings*. Springer, 62–79. doi:10.1007/978-3-319-29604-3_5

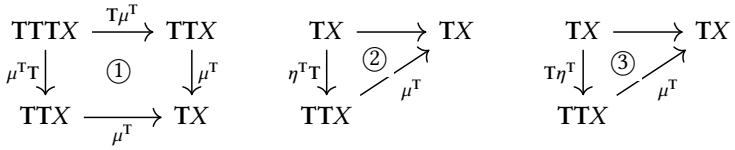
[48] Prakash Panangaden. 1999. The Category of Markov Kernels. *Electronic Notes in Theoretical Computer Science* 22 (Jan. 1999), 171–187. doi:10.1016/S1571-0661(05)80602-4

[49] Sungwoo Park. 2003. A calculus for probabilistic languages. In *Proceedings of TLDI'03: 2003 ACM SIGPLAN International Workshop on Types in Languages Design and Implementation, New Orleans, Louisiana, USA, January 18, 2003*, Zhong

Shao and Peter Lee (Eds.). ACM, 38–49. doi:10.1145/604174.604180
[50] Sungwoo Park, Frank Pfenning, and Sebastian Thrun. 2008. A probabilistic language based on sampling functions. *ACM Trans. Program. Lang. Syst.* 31, 1 (2008), 4:1–4:46. doi:10.1145/1452044.1452048
[51] Judea Pearl. 2009. *Causality.* Cambridge University Press.
[52] Robin Piedeleu. 2023. Basic Causal Inference via String Diagrams. *Blog post* (2023). Available at https://piedeleu.com/posts/diagrammatic-causal-inference/. Accessed on January 2026. Archived at https://web.archive.org/web/20241116235642/https://piedeleu.com/posts/diagrammatic-causal-inference/.
[53] Steve Selvin. 1975. Letters to the Editor. *The American Statistician* 29, 1 (1975), 67–71. doi:10.1080/00031305.1975.10479121
[54] Areeb Shah Mohammed. 2025. Partializations of Markov categories. arXiv:2509.05094 [math.CT] https://arxiv.org/abs/2509.05094
[55] Alexandra Silva and Ana Sokolova. 2011. Sound and Complete Axiomatization of Trace Semantics for Probabilistic Systems. In *Twenty-seventh Conference on the Mathematical Foundations of Programming Semantics, MFPS 2011, Pittsburgh, PA, USA, May 25-28, 2011 (Electronic Notes in Theoretical Computer Science, Vol. 276)*, Michael W. Mislove and Joël Ouaknine (Eds.). Elsevier, 291–311. doi:10.1016/J.ENTCS.2011.09.027
[56] Alex Simpson. 2017. Probability Sheaves and the Giry Monad. In *7th Conference on Algebra and Coalgebra in Computer Science, CALCO 2017, Ljubljana, Slovenia, June 12-16, 2017 (LIPIcs, Vol. 72)*, Filippo Bonchi and Barbara König (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 1:1–1:6. doi:10.4230/LIPICS.CALCO.2017.1
[57] Alex Simpson. 2024. Equivalence and Conditional Independence in Atomic Sheaf Logic. In *Proceedings of the 39th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2024, Tallinn, Estonia, July 8-11, 2024*, Pawel Sobocinski, Ugo Dal Lago, and Javier Esparza (Eds.). ACM, 70:1–70:14. doi:10.1145/3661814.3662132
[58] Ana Sokolova and Harald Woracek. 2018. Termination in Convex Sets of Distributions. *Log. Methods Comput. Sci.* 14, 4 (2018). doi:10.23638/LMCS-14(4:17)2018
[59] Eugene W Stark and Scott A Smolka. 2000. A complete axiom system for finite-state probabilistic processes. In *Proof, language, and interaction: essays in honour of Robin Milner*. 571–595.
[60] Sam Staton. 2017. Commutative semantics for probabilistic programming. In *European Symposium on Programming*. Springer, 855–879.
[61] Sam Staton. 2020. *Probabilistic Programs as Measures.* Cambridge University Press, 43–74.
[62] Sam Staton, Frank Wood, Hongseok Yang, Chris Heunen, and Ohad Kammar. 2016. Semantics for probabilistic programming: higher-order functions, continuous distributions, and soft constraints. In *2016 31st annual ACM/IEEE Symposium on Logic in Computer Science (LiCS)*. IEEE, 1–10.
[63] Dario Stein. 2025. Random Variables, Conditional Independence and Categories of Abstract Sample Spaces. In *40th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2025, Singapore, June 23-26, 2025*. IEEE, 472–484. doi:10.1109/LICS65433.2025.00042
[64] Ross Street. 2009. Weak distributive laws. *Theory and Applications of Categories* 22 (2009), 313–320.
[65] David Tolpin, Jan-Willem van de Meent, Hongseok Yang, and Frank D. Wood. 2016. Design and Implementation of Probabilistic Programming Language Anglican. In *Proceedings of the 28th Symposium on the Implementation and Application of Functional Programming Languages, IFL 2016, Leuven, Belgium, August 31 - September 2, 2016*, Tom Schrijvers (Ed.). ACM, 6:1–6:12. doi:10.1145/3064899.3064910
[66] Yun Chen Tsai, Kittiphon Phalakarn, S. Akshay, and Ichiro Hasuo. 2025. Chance and Mass Interpretations of Probabilities in Markov Decision Processes. In *36th International Conference on Concurrency Theory, CONCUR 2025, Aarhus, Denmark, August 26-29, 2025 (LIPIcs, Vol. 348)*, Patricia Bouyer and Jaco van de Pol (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 33:1–33:19. doi:10.4230/LIPICS.CONCUR.2025.33
[67] Matthijs Vákár, Ohad Kammar, and Sam Staton. 2019. A domain theory for statistical probabilistic programming. *Proceedings of the ACM on Programming Languages* 3, POPL (2019), 1–29.
[68] Daniele Varacca and Glynn Winskel. 2006. Distributing probability over non-determinism. *Math. Struct. Comput. Sci.* 16, 1 (2006), 87–113. doi:10.1017/S0960129505005074
[69] Marilyn vos Savant. [n. d.]. Parade 16: Ask Marilyn (Archived). https://web.archive.org/web/20130121183432/http://marilynvossavant.com/game-show-problem/. Accessed: 2013-01-21.
[70] Philip Wadler. 1992. The Essence of Functional Programming. In *Conference Record of the Nineteenth Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, Albuquerque, New Mexico, USA, January 19-22, 1992*, Ravi Sethi (Ed.). ACM Press, 1–14. doi:10.1145/143165.143169
[71] Mitchell Wand, Ryan Culpepper, Theophilos Giannakopoulos, and Andrew Cobb. 2018. Contextual equivalence for a probabilistic language with continuous random variables and recursion. *Proc. ACM Program. Lang.* 2, ICFP (2018), 87:1–87:30. doi:10.1145/3236782
[72] Yimu Yin and Jiji Zhang. 2022. Markov categories, causal theories, and the do-calculus. arXiv:2204.04821 [cs.AI]

[73] Maaike Zwart and Dan Marsden. 2022. No-Go Theorems for Distributive Laws. *Logical Methods in Computer Science* Volume 18, Issue 1, Article 13 (Jan 2022). doi:10.46298/lmcs-18(1:13)2022

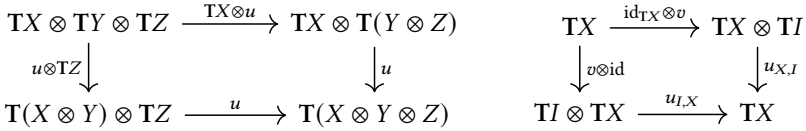
Definition .1 (Monad)**.** A *monad*, $(\mathbf{T}, \mu^{\mathbf{T}}, \eta^{\mathbf{T}})$, consists of an endofunctor, $\mathbf{T}\colon \mathbb{C} \to \mathbb{C}$, together with natural transformations, $\eta^{\mathbf{T}}\colon X \to \mathbf{T}X$ and $\mu^{\mathbf{T}}\colon \mathbf{TT}X \to \mathbf{T}X$, making the following three diagrams commute.



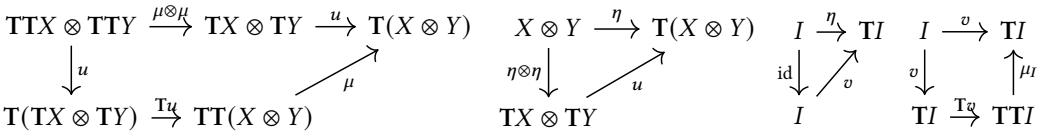
Definition .2 (Monoidal monad)**.** A *monoidal monad*, $(\mathbf{T}, \mu^{\mathbf{T}}, \eta^{\mathbf{T}}, u, v)$, on a monoidal category $(\mathbb{C}, \otimes, I)$ consists of a monad $(\mathbf{T}, \mu^{\mathbf{T}}, \eta^{\mathbf{T}})$, whose underlying functor is lax monoidal: i.e., there exist structural natural transformations,

$$u_{X,Y}\colon \mathbf{T}X \otimes \mathbf{T}Y \to \mathbf{T}(X \otimes Y) \;\text{ and }\; v\colon I \to \mathbf{T}I,$$

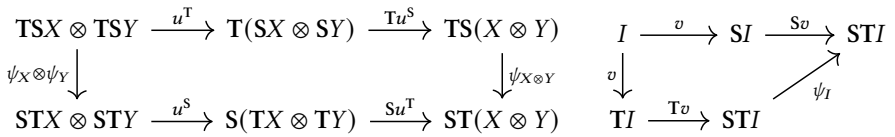
satisfying associativity, $(u_{X,Y} \otimes \mathrm{id}_Z) \fatsemi u_{X\otimes Y,Z} = (\mathrm{id}_X \otimes u_{Y,Z}) \fatsemi u_{X,Y\otimes Z}$, and unitality, $(\mathrm{id}_{\mathbf{T}X} \otimes v) \fatsemi u = \mathrm{id}_{\mathbf{T}X}$ and $(v \otimes \mathrm{id}_{\mathbf{T}X}) \fatsemi u = \mathrm{id}_{\mathbf{T}X}$. In other words, the following diagrams must commute.



Moreover, the unit and multiplication of the monad must be monoidal natural transformations, meaning that the following diagrams must commute.



Definition .3 (Monoidal distributive law)**.** A monoidal distributive law between monoidal monads, $(\mathbf{S}, \mu^{\mathbf{S}}, \eta^{\mathbf{S}}, u^{\mathbf{S}}, v^{\mathbf{S}})$ and $(\mathbf{T}, \mu^{\mathbf{T}}, \eta^{\mathbf{T}}, u^{\mathbf{T}}, v^{\mathbf{T}})$, is a distributive law, $\psi_X\colon \mathbf{TS}X \to \mathbf{ST}X$, whose transformation is monoidal, meaning that the two following diagrams must commute.



Theorem 2.10 (No-go result, cf. [58, Theorem 5.3]). *theorem]thm:no-go Black-hole semantics constitutes the only distributive law between the distribution monad and the maybe monad,* $\mathbf{DM} \to \mathbf{MD}$.

Proof. We will prove that any distributive law, $\psi\colon \mathbf{DM}X \to \mathbf{MD}X$, must coincide with black-hole semantics. We start by proving that, given any non-unit scalar $\lambda \in (0, 1)$, it must be the case that $\psi(\lambda\,[\langle x\rangle]) = \bot$ or $\psi(\lambda\,[\langle x\rangle]) = [x]$, for all $x \in X$. We first note that $\psi(\lambda\,[\langle *\rangle]) \in \mathbf{MD}1$ must either be $\psi(\lambda\,[\langle *\rangle]) = \langle [*]\rangle$ or $\psi(\lambda\,[\langle *\rangle]) = \bot$. By naturality over the function $(* \mapsto x)\colon 1 \to X$ that picks an element $x \in X$, we have $\psi(\lambda\,[\langle x\rangle]) = \mathbf{MD}(* \mapsto x)(\psi(\lambda\,[*]))$; and we conclude that either $\psi(\lambda\,[\langle x\rangle]) = \langle [x]\rangle$ or $\psi(\lambda\,[\langle x\rangle]) = \bot$.

Let us assume $\psi(\lambda\,[\langle x\rangle]) = \langle[x]\rangle$ and reach a contradiction with the D-multiplicativity axiom. Consider the distributions of subdistributions $d_1, d_2 \in \mathrm{DDM}X$ defined by

$$d_1 = \tfrac{1}{1+\lambda}\,[\lambda\,[\langle x\rangle] + (1-\lambda)\,[\bot]] + \tfrac{\lambda}{1+\lambda}\,[[\langle y\rangle]]\,;\text{ and}$$
$$d_2 = \tfrac{1}{1+\lambda}\,[\lambda\,[\langle y\rangle] + (1-\lambda)\,[\bot]] + \tfrac{\lambda}{1+\lambda}\,[[\langle x\rangle]]\,.$$

Applying one side of the D-multiplicativity axiom yields the same result on both distributions; it follows that applying the other side of the axiom should also coincide.

$$\psi \circ \mu\mathrm{M}(d_1) = \psi\left(\tfrac{\lambda}{1+\lambda}\,[\langle x\rangle] + \tfrac{\lambda}{1+\lambda}\,[\langle y\rangle] + \tfrac{1-\lambda}{1+\lambda}\,[\bot]\right) = \psi \circ \mu\mathrm{M}(d_2).$$

However, we can *(i)* take both distributions of subdistributions; *(i)* normalize internally, using the assumption that $\psi(\rho\,[x]) = [x]$ and that $\psi([x]) = 1$ by the D-unitality axiom; *(ii)* normalize externally using that $\psi(\sum_{i=0}^{n}\lambda_i\,[x_i]) = \sum_{i=0}^{n}[x_i]$ whenever $\sum_{i=0}^{n}\lambda_i = 1$ by the M-unitality axiom; and *(iii)* multiply, obtaining two different results.

$$\begin{aligned}
\mathrm{M}\mu \circ \psi\mathrm{D} \circ \mathrm{D}\psi\,(d_1) &\overset{(i)}{=} \\
\mathrm{M}\mu \circ \psi\mathrm{D}\left(\tfrac{1}{1+\lambda}\,[\langle[x]\rangle] + \tfrac{\lambda}{1+\lambda}\,[\langle[y]\rangle]\right) &\overset{(ii)}{=} \\
\mathrm{M}\mu\left(\left\langle\tfrac{1}{1+\lambda}\,[[x]] + \tfrac{\lambda}{1+\lambda}\,[[y]]\right\rangle\right) &\overset{(iii)}{=} \\
\left\langle\tfrac{1}{1+\lambda}\,[x] + \tfrac{\lambda}{1+\lambda}\,[y]\right\rangle.
\end{aligned}
\qquad
\begin{aligned}
\mathrm{M}\mu \circ \psi\mathrm{D} \circ \mathrm{D}\psi\,(d_2) &\overset{(i)}{=} \\
\mathrm{M}\mu \circ \psi\mathrm{D}\left(\tfrac{1}{1+\lambda}\,[\langle[x]\rangle] + \tfrac{\lambda}{1+\lambda}\,[\langle[y]\rangle]\right) &\overset{(ii)}{=} \\
\mathrm{M}\mu\left(\left\langle\tfrac{\lambda}{1+\lambda}\,[[x]] + \tfrac{1}{1+\lambda}\,[[y]]\right\rangle\right) &\overset{(iii)}{=} \\
\left\langle\tfrac{\lambda}{1+\lambda}\,[x] + \tfrac{1}{1+\lambda}\,[y]\right\rangle.
\end{aligned}$$

Because $\frac{\lambda}{1+\lambda} \neq \frac{1}{1+\lambda}$ unless $\lambda = 1$, these two normalized distributions are different, and we must conclude $\psi(\lambda\,[\langle x\rangle]) = \bot$.

Finally, let us prove that the image of any non-full subdistribution must be zero. Consider an arbitrary non-zero and non-full subdistribution, $\sum_i \lambda_i\,[\langle x_i\rangle] + (1-\lambda)\,[\bot]$, where $\lambda = \sum_i \lambda_i \neq 1$. We reason as follows.

$$\begin{aligned}
&\psi\left(\textstyle\sum_i \lambda_i\,[\langle x_i\rangle] + (1-\lambda)\,[\bot]\right) && \overset{(i)}{=} \\
&\psi \circ \mu^{\mathrm{D}}\mathrm{M}\left(\lambda\,\left[\textstyle\sum_i \tfrac{\lambda_i}{\lambda}\,[\langle x_i\rangle]\right] + (1-\lambda)\,[[\bot]]\right) && \overset{(ii)}{=} \\
&\mathrm{M}\mu^{\mathrm{D}} \circ \psi\mathrm{D} \circ \mathrm{D}\psi\left(\lambda\,\left[\textstyle\sum_i \tfrac{\lambda_i}{\lambda}\,[\langle x_i\rangle]\right] + (1-\lambda)\,[[\bot]]\right) && \overset{(iii)}{=} \\
&\mathrm{M}\mu^{\mathrm{D}} \circ \psi\mathrm{D}\left(\lambda\,\left[\left\langle\textstyle\sum_i \tfrac{\lambda_i}{\lambda}\,[x_i]\right\rangle\right] + (1-\lambda)\,[\bot]\right) && \overset{(iv)}{=} \\
&\mathrm{M}\mu^{\mathrm{D}}\left(\bot\right) && \overset{(v)}{=} \\
&\bot.
\end{aligned}$$

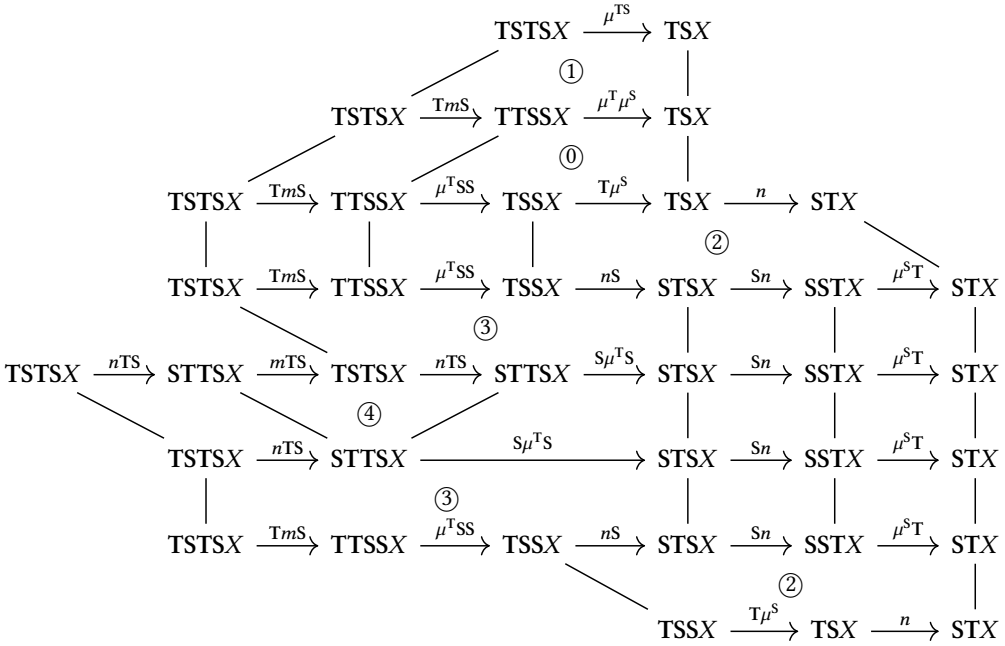
We first find *(i)* a distribution that multiplies to the arbitrary non-zero and non-full subdistribution, *(ii)* apply the D-multiplicativity axiom, *(iii)* use the fact that ψ leaves full distributions and failure invariant; we *(iv)* use the previous result that states that ψ sends every non-full distribution to failure, and *(v)* we conclude by definition of the maybe monad.

We conclude that the distributive law, when applied to any non-full and non-zero subdistribution must return zero; and that, by unitality, when applied to a full or zero subdistribution, it must return these unchanged. We conclude it coincides with *black-hole semantics*. □

Lemma 3.5 (Renormalization). *Any sesquilaw,* $(\mathrm{S}, \mathrm{T}, m, n)$*, induces an idempotent,* $k = (n \fatsemi m) : \mathrm{TS} \to \mathrm{TS}$*. This idempotent is left-absorptive, meaning that the following diagram commutes.*

$$\begin{array}{ccccc}
\mathrm{TSTS}X & \xrightarrow{\mu_m^{\mathrm{TS}}} & \mathrm{TS}X & \xrightarrow{n} & \mathrm{ST}X \\
{\scriptstyle n\mathrm{TS}}\downarrow & & & & \uparrow{\scriptstyle n} \\
\mathrm{STTS}X & \xrightarrow{m\mathrm{TS}} & \mathrm{TSTS}X & \xrightarrow{\mu_m^{\mathrm{TS}}} & \mathrm{TS}X
\end{array}$$

Proof. We employ commutative diagrams. We use ⓪ naturality, ① the definition of the monad arising from a distributive law, ② the S-multiplicativity of the quasidistributive law n, ③ the first sesquilaw axiom, and ④ the second sesquilaw axiom.


TSTSX
TSTSX
TSTSX

On the right hand side, we use the normalization and monad multiplication.

$$\begin{aligned}\mathsf{p}(\mu(\psi(-^\bullet)))(x) &= \frac{\mu(\psi(-^\bullet))(x)}{\bigvee_{y\in X}\mu(\psi(-^\bullet))(y)}\oslash = \frac{\bigvee_{\alpha\in \mathrm{D}X}\psi(\alpha^\bullet)\wedge\alpha(x)}{\bigvee_{y\in X}\bigvee_{\alpha\in \mathrm{R}X}\psi(\alpha^\bullet)\wedge\alpha(y)}\oslash \\ &= \frac{\bigvee_{\alpha\in \mathrm{D}X}\psi(\alpha^\bullet)\wedge\alpha(x)}{\bigvee_{\alpha\in \mathrm{D}X}\psi(\alpha^\bullet)\wedge\bigvee_{y\in X}\psi(y)}\oslash = \frac{\bigvee_{\alpha\in \mathrm{D}X}\psi(\alpha^\bullet)\wedge\alpha(x)}{\bigvee_{\alpha\in \mathrm{D}X}\psi(\alpha^\bullet)}\oslash.\end{aligned}$$

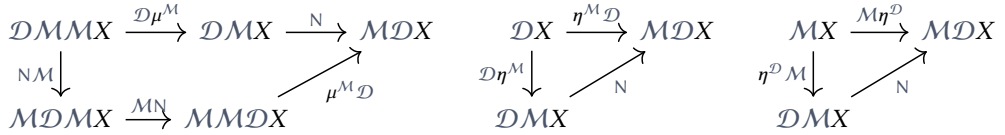
The last step uses that all $\alpha \in \mathrm{R}X$ are non-null predicates: $\bigvee_{y\in X}\alpha(y)$ must be exactly 1. The rest of the proof is analogous. □

Theorem 3.13. *Normalization in standard Borel spaces,* $\mathsf{N}(-)\colon \mathcal{DM}X \to \mathcal{MD}X$*, yields a sesquilaw.*

Proof. During this proof, we use zero-handling division ($\oslash$), defined by $x\oslash 0 = 0$ and $x\oslash y = x/y$ whenever $y \neq 0$. Let us first prove that normalization is monoidal. Consider two measures, $\nu_1 \in \mathcal{D}X$ and $\nu_2 \in \mathcal{D}Y$. We use *(i)* the definition of normalization and the tensor of measures, *(ii)* the definition of the indicator function, *(iii)* linearity of the Lebesgue integral, *(iv)* linearity of the Lebesgue integral, and *(v)* the definition of the tensor of measures.

$$\begin{aligned}\mathsf{N}(\nu_1\otimes\nu_2)(U) &\overset{(i)}{=} \frac{\int_{y\in Y}\left(\int_{x\in X}\xi_U(x,y)\cdot\nu_1(\mathrm{d}x)\right)\cdot\nu_2(\mathrm{d}y)}{\int_{y\in Y}\left(\int_{x\in X}\xi_{X\times Y}(x,y)\cdot\nu_1(\mathrm{d}x)\right)\cdot\nu_2(\mathrm{d}y)}\oslash \\ &\overset{(ii)}{=} \frac{\int_{y\in Y}\left(\int_{x\in X}\xi_U(x,y)\cdot\nu_1(\mathrm{d}x)\right)\cdot\nu_2(\mathrm{d}y)}{\int_{y\in Y}\left(\int_{x\in X}1\cdot\nu_1(\mathrm{d}x)\right)\cdot\nu_2(\mathrm{d}y)}\oslash \\ &\overset{(iii)}{=} \frac{\int_{y\in Y}\left(\int_{x\in X}\xi_U(x,y)\cdot\nu_1(\mathrm{d}x)\right)\cdot\nu_2(\mathrm{d}y)}{\nu_1(X)\cdot\nu_2(Y)}\oslash \\ &\overset{(iv)}{=} \int_{y\in Y}\left(\int_{x\in X}\xi_U(x,y)\cdot\frac{\nu_1(\mathrm{d}x)}{\nu_1(X)}\oslash\right)\cdot\frac{\nu_2(\mathrm{d}y)}{\nu_2(Y)}\oslash \overset{(v)}{=} (\mathsf{N}(\nu_1)\otimes\mathsf{N}(\nu_2))(U).\end{aligned}$$

Let us now prove that normalization determines a quasidistributive law. We must prove that the following three diagrams commute.



For $\mathcal{M}$-multiplicativity, note that an element $d \in \mathcal{DMM}X$ is a measure over $X + \{\bot\} + \{\bot'\}$. Let us reason by cases. If $d(\{\bot,\bot'\}) = 1$, then $\mathsf{N}(\mu^{\mathcal{M}}(d)) = \bot = \mu^{\mathcal{M}}(\mathsf{N}(\mathsf{N}(d)))$, either because $d(\bot') = 1$, or because $d(\bot') \neq 1$ but then $d(\bot) > 0$ and $\mathsf{N}(d)(\bot) = 1$. Assume, thus, that $d(\{\bot,\bot'\}) \neq 1$. We then need to prove that normalizing both separately is the same as normalizing after identifying both.

$$\mu^{\mathcal{M}}(\mathsf{N}(\mathsf{N}(d)))(U) = \frac{\mathsf{N}(d)(U)}{\mathsf{N}(d)(X)}\oslash = \frac{\frac{d(U)}{d(X+\{\bot\})}\oslash}{\frac{d(X)}{d(X+\{\bot\})}\oslash}\oslash = \frac{d(U)}{d(X)}\oslash = \mathsf{n}(\mu^{\mathrm{M}}(d)).$$

For $\mathcal{M}$-unitality, we must check that the normalization of a normalized distribution is itself: by $\eta^{\mathcal{M}}(d)(U) = d(U)$, we conclude that $\eta^{\mathrm{M}}(d)(X) = d(X) = 1$, and thus $\mathsf{N}(\eta^{\mathcal{M}}(d)) = \eta^{\mathcal{M}}(d)$. For $\mathcal{D}$-unitality, we reason by cases on $\mathcal{M}X$. We check that $\mathsf{N}(\eta^{\mathrm{D}}(\bot)) = \mathsf{N}([\bot]) = \bot$ and that $\mathsf{N}(\eta^{\mathrm{D}}(x)) = \mathsf{N}([x]) = x$.

Let us now prove that normalization and subdistributions form a sesquilaw.

We first need to prove that normalization and the inclusion into subdistributions are a section-retraction pair. Given any normalized distribution, $\nu \in \mathcal{M}\mathcal{D}X$, the measure of the whole set must either be zero or one, $\nu(X) = 1$ or $\nu(X) = 0$. If it is zero, it follows that the measure of any subset, $U \subseteq X$, must also be zero, $\nu(U) = 0$; as a consequence, in any of the two cases,

$$\mathsf{N}(\nu)(U) = \frac{\nu(U)}{\nu(X)}\oslash = \nu(U).$$

Let us prove that the second formulation of the axiom of distributive sesquilaws holds.

$$\begin{array}{ccccc} \mathcal{D}\mathcal{M}\mathcal{D}X & \xrightarrow{\mathsf{N}\mathcal{D}} & \mathcal{M}\mathcal{D}\mathcal{D}X & \xrightarrow{\mathcal{M}\mu} & \mathcal{M}\mathcal{D}X \\ {\scriptstyle \mathcal{D}m}\downarrow & & & \nearrow_{\mathsf{N}} & \\ \mathcal{D}\mathcal{D}\mathcal{M}X & \xrightarrow{\mu\mathcal{M}} & \mathcal{D}\mathcal{M}X & & \end{array}$$

Let $\nu \in \mathcal{D}\mathcal{M}\mathcal{D}X$ be a subdistribution of distributions. We must prove that normalizing and flattening the distributions is the same as, while regarding the distributions as subdistributions, flattening and then normalizing. In other words, we seek to prove

$$\mathcal{M}\mu(\mathsf{N}(\nu)) = \mathsf{N}(\mu(\nu(-^\bullet))).$$

On the left hand side, we use *(i)* the definition of monad multiplication for the Giry monad, and *(ii)* the definition of normalization.

$$\mathcal{M}\mu(\mathsf{N}(\nu(-^\bullet)))(U) \overset{(i)}{=} \int_{\alpha\in\mathcal{D}X} \alpha(U)\cdot \mathsf{N}(\nu(-^\bullet))(\mathrm{d}\alpha) \overset{(ii)}{=} \int_{\alpha\in\mathcal{D}X} \alpha(U)\cdot \frac{\nu(\mathrm{d}\alpha)}{\nu(\mathcal{D}X)}\oslash.$$

On the right hand side, we use *(i)* the definition of normalization, *(ii)* the definition of monad multiplication for the Giry monad and, finally, *(iii)* that the $\alpha \in \mathcal{D}X$ elements are full distributions, meaning that $\alpha(X) = 1$.

$$\mathsf{N}(\mu(\nu))(U) \overset{(i)}{=} \frac{\mu(\nu)(U)}{\mu(\nu)(X)}\oslash \overset{(ii)}{=} \frac{\int_{\alpha\in\mathcal{D}X}\alpha(U)\cdot\nu(\mathrm{d}\alpha)}{\int_{\alpha\in\mathcal{D}X}\alpha(X)\cdot\nu(\mathrm{d}\alpha)}\oslash \overset{(iii)}{=} \frac{\int_{\alpha\in\mathcal{D}X}\alpha(U)\cdot\nu(\mathrm{d}\alpha)}{\nu(\mathcal{D}X)}\oslash$$

Lastly, we may divide by cases: whether $\nu(\mathcal{D}X) = 0$ or $\nu(\mathcal{D}X) \neq 0$, we may use linearity of the Lebesgue integral to equate the two sides. □

Proposition 3.17 (Support). *Support,* $\mathrm{supp}_X : \mathbf{D}X \to \mathbf{R}X$*, is a sesquilaw morphism — with* $\mathbf{F} = \mathrm{Id}$ *the identity functor — between normalization and postselection.*

Proposition 4.2. *Any quasidistributive law,* $\psi : \mathbf{TS}X \to \mathbf{ST}X$*, induces a magmad,* $(\mathbf{ST}, \mu^{\mathbf{ST}}_\psi, \eta^{\mathbf{ST}}_\psi)$ *defined by* $\mu^{\mathbf{ST}}_\psi = \mathbf{S}\psi\mathbf{T} \fatsemi \mu^{\mathbf{S}}\mu^{\mathbf{T}}$ *and* $\eta^{\mathbf{ST}}_\psi = \eta^{\mathbf{S}}\eta^{\mathbf{T}}$.

Proof. The unit of the composite non-associative monad is $\eta = \eta^{\mathbf{S}}\eta^{\mathbf{T}}$; the multiplication is $\mu = \mathbf{S}\psi\mathbf{T} \fatsemi \mu^{\mathbf{S}}\mu^{\mathbf{T}}$. From unitality of the quasidistributive law, unitality of the non-associative monad follows. Whenever the quasidistributive law is moreover monoidal, the unit and multiplication of the monad become monoidal by construction. □

Theorem 4.5. *Any sesquilaw,* $(\mathbf{S}, \mathbf{T}, m, n)$*, induces a right action of its monad,* $\mathbf{TS}$*, on its magmad,* $\mathbf{ST}$*: i.e., a natural transformation,* $u : \mathbf{STTS}X \to \mathbf{ST}X$*, making the following two diagrams commute.*

$$\begin{array}{ccc} \mathbf{ST}X & \xrightarrow{\eta^{\mathbf{TS}}} & \mathbf{STTS}X \\ & \searrow^{\mathrm{id}} & \downarrow{\scriptstyle u} \\ & & \mathbf{ST}X \end{array} \qquad \begin{array}{ccc} \mathbf{STTSTS}X & \xrightarrow{\mathbf{ST}\mu^{\mathbf{TS}}} & \mathbf{STTS}X \\ {\scriptstyle u\mathbf{TS}}\downarrow & & \downarrow{\scriptstyle u} \\ \mathbf{STTS}X & \xrightarrow{u} & \mathbf{ST}X \end{array}$$

This action is defined by either side of the following commutative diagram.

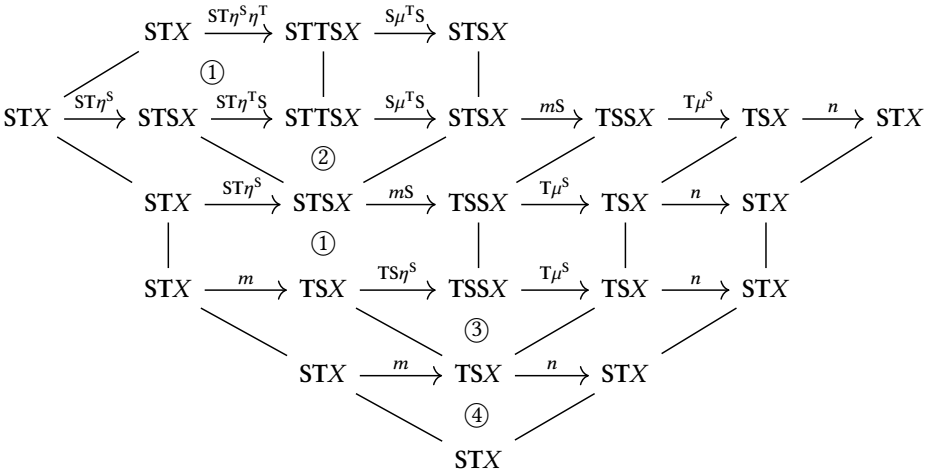
$$\begin{array}{ccccccc}
\mathbf{STTS}X & \xrightarrow{\mathbf{S}\mu^{\mathbf{T}}\mathbf{S}} & \mathbf{STS}X & \xrightarrow{m\mathbf{S}} & \mathbf{TSS}X & & \\
{\scriptstyle m\mathbf{TS}}\downarrow & & & & \downarrow{\scriptstyle \mathbf{T}\mu^{\mathbf{S}}} & & \\
\mathbf{TSTS}X & \xrightarrow{\mathbf{T}m\mathbf{S}} & \mathbf{TTSS}X & \xrightarrow{\mu^{\mathbf{T}}\mu^{\mathbf{S}}} & \mathbf{TS}X & \xrightarrow{n} & \mathbf{ST}X
\end{array}$$

Proof. Let us use commutative diagrams for this proof. We will show that any sesquilaw, $(m, n, \mathbf{S}, \mathbf{T})$, induces an action of th monad $\mathbf{TS}$ into the almost monad $\mathbf{ST}$. We define the action, $u\colon \mathbf{STTS} \to \mathbf{ST}$, by string diagrams, as any of the two following equivalent definitions.

$$\begin{array}{ccccccc}
\mathbf{STTS}X & \xrightarrow{\mathbf{S}\mu^{\mathbf{T}}\mathbf{S}} & \mathbf{STS}X & \xrightarrow{m\mathbf{S}} & \mathbf{TSS}X & & \\
{\scriptstyle m\mathbf{TS}}\downarrow & & & & \downarrow{\scriptstyle \mathbf{T}\mu^{\mathbf{S}}} & & \\
\mathbf{TSTS}X & \xrightarrow{\mathbf{T}m\mathbf{S}} & \mathbf{TTSS}X & \xrightarrow{\mu^{\mathbf{T}}\mu^{\mathbf{S}}} & \mathbf{TS}X & \xrightarrow{n} & \mathbf{ST}X
\end{array}$$

Let us prove multiplicativity (in Figure 4). We use ⓪ naturality, ① S-multiplicativity of the quasidistributive law n, ② the sesquilaw axiom, ③ that sesquilaws are inverses, ④ S-multiplicativity of the distributive law m, ⑤ associativity of S, ⑥ T-multiplicativity of the distributive law m, and ⑦ associativity of the monad T.

The proof of unitality is more direct: it uses ① naturality, ② unitality of the monad T, ③ unitality of the monad S, and ④ that sesquilaws are inverses.



With these two axioms, we have built an action. In the monoidal case, we repeat the exact same proof just considering that all natural transformations are monoidal natural transformations. □

Proposition 4.8. *proposition]prop:magmoid-normalized-stochastic-kernels* Normalized stochastic kernels *between sets,* $X \to \mathbf{MD}Y$*, form a magmoid,* **Norm**. *Given* $f\colon X \to \mathbf{MD}Y$*, we write* $f(x;y)$ *for the coefficient of* $y \in Y$ *in the normalized distribution* $f(x)$*. Composition of two morphisms,* $f\colon X \to \mathbf{MD}Y$ *and* $g\colon Y \to \mathbf{MD}Z$*, is defined by*

$$(f \fatsemi g)(x;z) = \frac{\sum_{v\in Y} f(x;v)\cdot g(v;z)}{\sum_{v\in Y}\sum_{w\in Z} f(x;v)\cdot g(v;w)}, \text{ when } \sum_{v,w} f(x;v)\cdot g(v;w) \neq 0; \text{ and } 0 \text{ otherwise.}$$

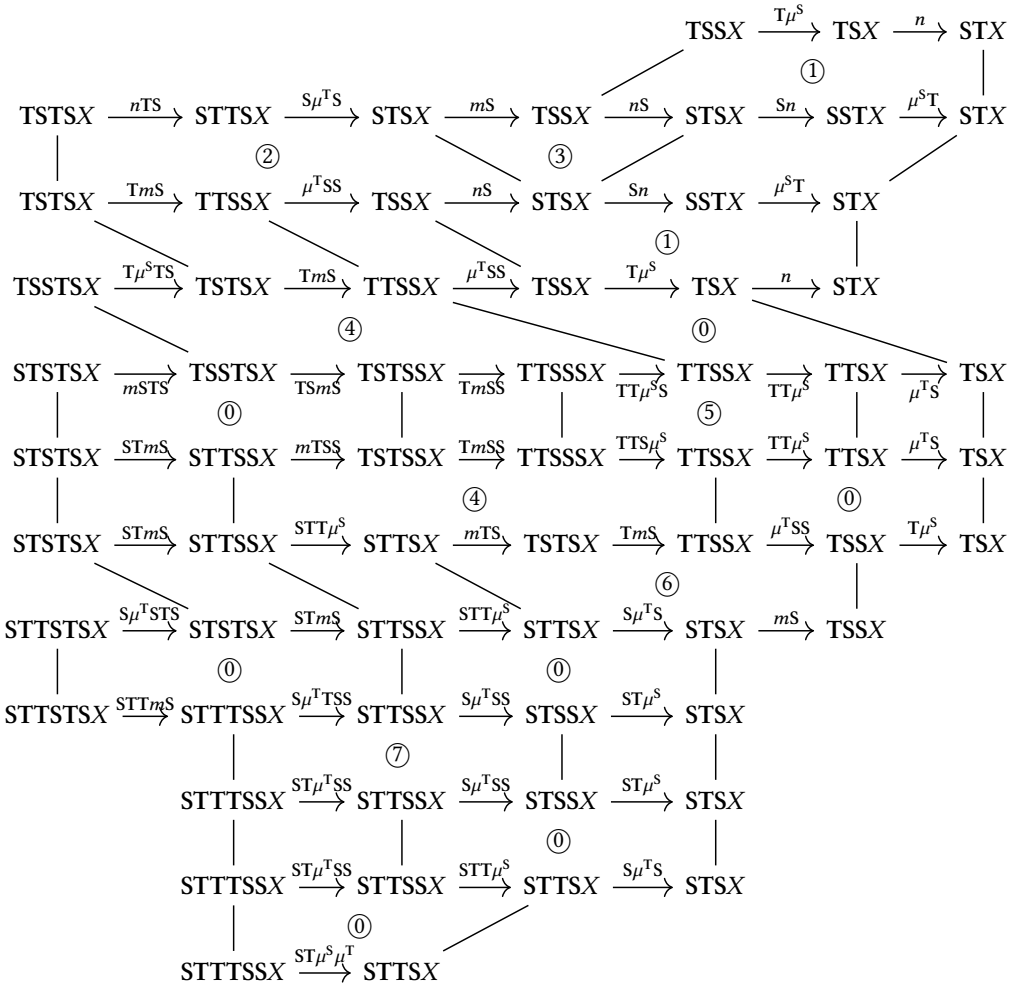


Fig. 4. Multiplicativity for the right action of a sesquilaw.

Proof. We define $\mathrm{id}(x;x') = [x = x']$. Let us check that it is right unital. Let $f\colon X \to \mathbf{M}\mathbf{D}Y$ be a normalized kernel.

$$(f \fatsemi \mathrm{id})(x;y) = \frac{\sum_{y'} f(x;y') \cdot [y' = y]}{\sum_{y',y} f(x;y') \cdot [y' = y]}\oslash = \frac{f(x;y)}{\sum_y f(x;y)}\oslash = f(x;y).$$

On the last step, either $\sum_{y'} f(x;y') = 0$, which implies $f(x;y) = 0$, or $\sum_{y'} f(x;y') = 1$, which simplifies the division. Left unitality is analogous. □

Proposition 4.10 (Kleisli magmoids). *proposition]prop:kleisli-magmoids Any magmad,* $(\mathbf{N}, \mu^{\mathbf{N}}, \eta^{\mathbf{N}})$ *over a category* $\mathbb{C}$ *with composition operation* $(;)$*, induces a magmoid,* $\mathbb{K}(\mathbf{N})$*, with its same objects, with morphisms* $\mathbb{K}(\mathbf{N})(X;Y) = \mathbb{C}(X;\mathbf{N}Y)$*, with identities* $\mathrm{id} = \eta^{\mathbf{N}}$*, and with compositions,* $(f \fatsemi g) = f;\mathbf{N}g;\mu^{\mathbf{N}}$.

Proposition 4.12 (Sesquilaw magmoids). *Any sesquilaw,* $(\mathbf{S}, \mathbf{T}, m, n)$*, induces a Kleisli magmoid where associativity is satisfied,* $f \fatsemi (g \fatsemi h) = (f \fatsemi g) \fatsemi h$*, if either*

(1) $f = u;\mathbf{S}\eta^{\mathbf{T}}$ *for some* $u\colon X \to \mathbf{S}Y$*;*

(2) $g = v \,;\, \mathrm{S}\eta^{\mathrm{T}}$ *for some* $v \colon Y \to \mathrm{S}Z$*; or*
(3) $h = w \,;\, \eta^{\mathrm{S}}\mathrm{T}$ *for some* $w \colon Z \to \mathrm{T}W$.

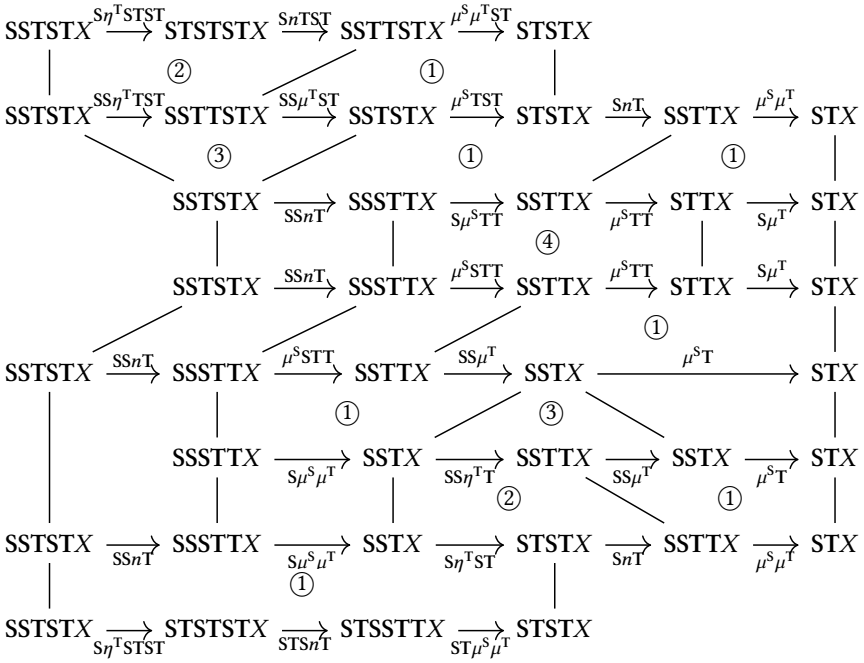
Proof. Proving these equations for the case of normalized kernels is straightforward — we only need to check that the resulting expressions are equal — but difficult to repeat in the general case. The general case requires diagram chasing. These three cases reduce, respectively, to the commutativity of the following three diagrams.

$$\begin{array}{ccccc}
\mathrm{SSTST}X & \xrightarrow{\mathrm{S}\eta^{\mathrm{T}}\mathrm{STST}} & \mathrm{STSTST}X & \xrightarrow{\mu^{\mathrm{ST}}\mathrm{ST}} & \mathrm{STST}X \\
{\scriptstyle \mathrm{S}\eta^{\mathrm{T}}\mathrm{STST}}\downarrow & & & & \downarrow{\scriptstyle \mu^{\mathrm{ST}}} \\
\mathrm{STSTST}X & \xrightarrow[\mathrm{ST}\mu^{\mathrm{ST}}]{} & \mathrm{STST}X & \xrightarrow[\mu^{\mathrm{ST}}]{} & \mathrm{ST}X
\end{array}$$

$$\begin{array}{ccccc}
\mathrm{STSST}X & \xrightarrow{\mathrm{STS}\eta^{\mathrm{T}}\mathrm{ST}} & \mathrm{STSTST}X & \xrightarrow{\mu^{\mathrm{ST}}\mathrm{ST}} & \mathrm{STST}X \\
{\scriptstyle \mathrm{STS}\eta^{\mathrm{T}}\mathrm{ST}}\downarrow & & & & \downarrow{\scriptstyle \mu^{\mathrm{ST}}} \\
\mathrm{STSTST}X & \xrightarrow[\mathrm{ST}\mu^{\mathrm{ST}}]{} & \mathrm{STST}X & \xrightarrow[\mu^{\mathrm{ST}}]{} & \mathrm{ST}X
\end{array}$$

$$\begin{array}{ccccc}
\mathrm{STSTT}X & \xrightarrow{\mathrm{STST}\eta^{\mathrm{S}}\mathrm{T}} & \mathrm{STSTST}X & \xrightarrow{\mu^{\mathrm{ST}}\mathrm{ST}} & \mathrm{STST}X \\
{\scriptstyle \mathrm{STST}\eta^{\mathrm{S}}\mathrm{T}}\downarrow & & & & \downarrow{\scriptstyle \mu^{\mathrm{ST}}} \\
\mathrm{STSTST}X & \xrightarrow[\mathrm{ST}\mu^{\mathrm{ST}}]{} & \mathrm{STST}X & \xrightarrow[\mu^{\mathrm{ST}}]{} & \mathrm{ST}X
\end{array}$$

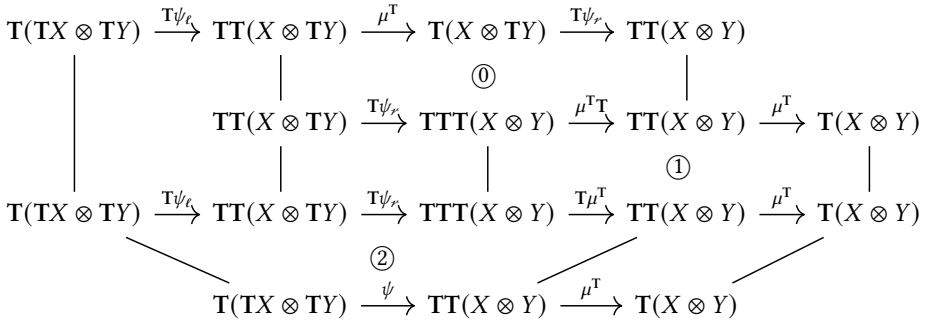
Let us consider the first case. We employ ① naturality, ② unitality of the quasidistributive law n, ③ unitality of the monad $\mathbf{T}$, and ④ associativity of the monad $\mathbf{S}$.



The other two cases follow a similar reasoning. □

Proposition 5.7. *proposition]prop:monad-left-exchange A monad is commutative if and only if it is left-exchanging. proposition]rem:counterexample-left-exchange Not every commutative magmad is a left-exchanging magmad.*

Proof. We employ ⓪ naturality, ① the associativity of the monad, and ② the definition of the joint strength.



Note how the proof employs the associativity of the monad. To prove that associativity is required, consider the *average-writer magmad*, $(\mathbf{Q}, \mu^{\mathbf{Q}}, \eta^{\mathbf{Q}})$, given by the functor $\mathbf{Q}X = X \times \mathbb{MQ}^+$, with the multiplication defined by $\mu^{\mathbf{Q}}(x, q_1, q_2) = (x, (q_1 + q_2)/2)$ — with $\mu^{\mathbf{Q}}(x, \bot, q_2) = (x, q_2)$ and $\mu^{\mathbf{Q}}(x, q_1, \bot) = (x, q_1)$ — and the unit defined by $\eta^{\mathbf{Q}}(x) = (x, \bot)$. This is a commutative magmad but not a left-exchanging magmad. □

Theorem 5.17 (Front-door adjustment formula [51]). *Given a magmad, $(\mathbf{N}, \mu^{\mathbf{N}}, \eta^{\mathbf{N}})$, satisfying the principles in Figure 3, let an expression, $P \in \mathbf{N}(X \times Y \times Z)$, admit the following factorization into total morphisms — $G \colon 1 \to \mathbf{N}U$, $S \colon U \to \mathbf{N}X$, $T \colon X \to \mathbf{N}Y$, and $C \colon Y \times U \to \mathbf{N}Z$ — where, moreover, both T and $(G \fatsemi S)$ have full support.*

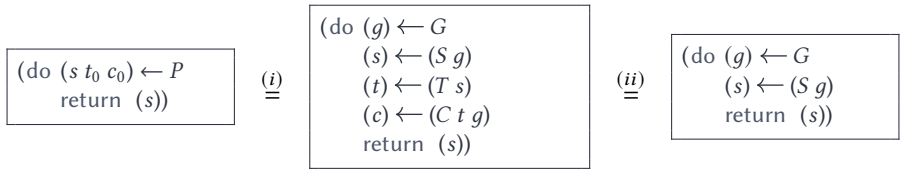
```
(do (s t c) ← P
    return (s t c))
```
=
```
(do (g) ← G
    (s) ← (S g)
    (t) ← (T s)
    (c) ← (C t g)
    return (s t c))
```

Then, the following equation holds. We define (frontDoor P a) *as its right-hand side.*

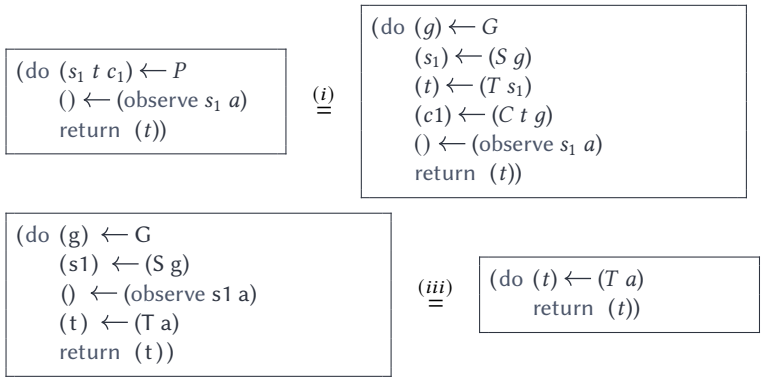
```
(do (g) ← G
    (t) ← (T a)
    (c) ← (C t g)
    return (c))
```
=
```
(do (s t0 c0) ← P
    (t) ← (do (s1 t c1) ← P
              () ← (observe s1 a)
              return (t))
    (c) ← (do (s2 t2 c) ← P
              () ← (observe s2 s)
              () ← (observe t2 t)
              return (c))
    return (c))
```

In other words, an intervention setting s to a can be derived as a function of the observational data.

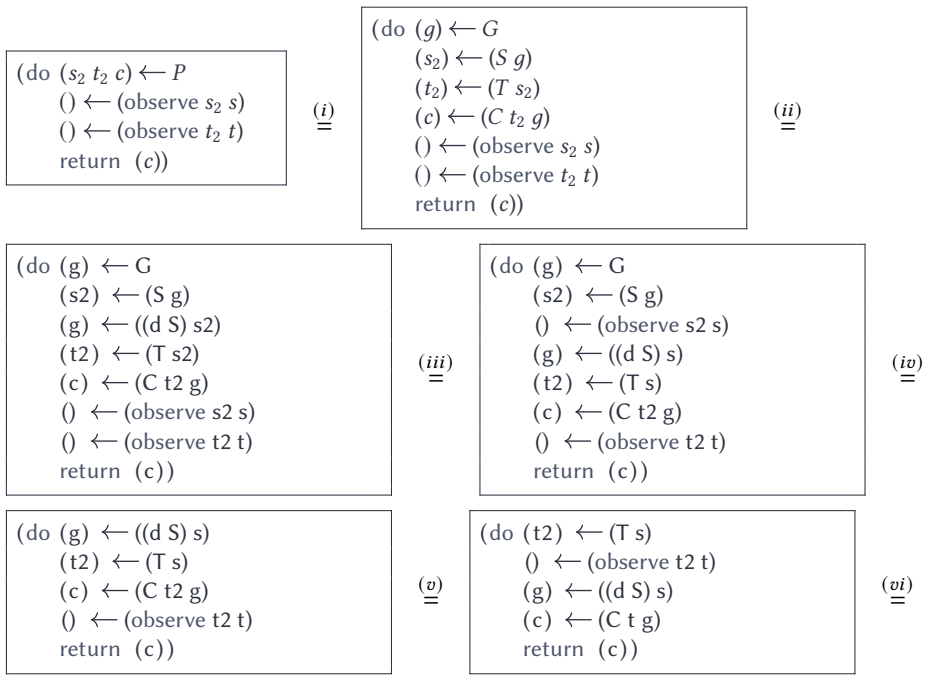
Proof. Let us simplify each one of the three factors. For the first factor, we use *(i)* the hypothesis, and *(ii)* totality of C.



For the second factor computes the marginal of a conditional distribution. We use *(i)* the hypothesis; *(ii)* totality of c and Frobenius; and *(iii)* that $(g \fatsemi s)$ has full support.



The third factor can be simplified as follows. We use *(i)* the hypothesis, *(ii)* the disintegration axiom, *(iii)* the Frobenius axiom and exchange, *(iv)* that $(g \fatsemi s)$ has full support, *(v)* exchange, and *(vi)* that (t) has full support.



```
(do (g) ← (∂S s)
    (c) ← (C t g)
    return (c))
```

Finally, let us address the main claim. We *(i)* substitute according to the previous three equations, *(ii)* use the totality of g and c, and *(iii)* apply the disintegration axiom.

```
(do (s t₀ c₀) ← P
    (t) ← (do (s₁ t c₁) ← P
              () ← (observe s₁ a)
              return (t))
    (c) ← (do (s₂ t₂ c) ← P
              () ← (observe s₂ s)
              () ← (observe t₂ t)
              return (c))
    return (c))
```

$\overset{(i)}{=}$

```
(do
  (t) ← (T a)
  (g) ← G
  (s) ← (S g)
  (c) ← (do
    (g) ← ((d S) s)
    (c) ← (C t g)
    return (c))
  return (c))
```

$\overset{(ii)}{=}$

```
(do (t) ← (T a)
    (g) ← G
    (s) ← (S g)
    (g) ← ((d S) s)
    (c) ← (C t g)
    return (c))
```

$\overset{(iii)}{=}$

```
(do (g) ← G
    (t) ← (T a)
    (c) ← (C t g)
    return (c))
```

This concludes the proof: the intervention can be written exclusively in terms of P. □

Theorem 5.18 (Back-door adjustment formula [51]). *Given a magmad,* $(\mathbf{N}, \mu^{\mathbf{N}}, \eta^{\mathbf{N}})$*, satisfying the principles in Figure 3, let an expression,* $P \in \mathbf{N}(U \times X \times Y)$*, admit the following factorization into total morphisms — where, moreover, F has full support.*

```
(do (u x y) ← P
    return (u x y))
```

$\overset{(i)}{=}$

```
(do (u) ← F
    (x) ← (G u)
    (y) ← (H x u)
    return (u x y))
```

Then, the following equation holds. We define (backDoor P a) *as the right-hand side.*

```
(do (u) ← F
    (y) ← (H a u)
    return (y))
```

$\overset{(ii)}{=}$

```
(do (u₀ x₀ y₀) ← P
    (y) ← (do (u x y) ← P
              () ← (observe u u₀)
              () ← (observe x a)
              return (y))
    return (y))
```

Proof. Let us first simplify the uppermost part of the diagram. We use *(i)* the hypothesis, and *(ii)* the totality of G and H.

```
(do (u₀ x₀ y₀) ← P
    return (u₀))
```

$\overset{(i)}{=}$

```
(do (u₀) ← F
    (x) ← (G x)
    (y) ← (H x u)
    return (u₀ x y))
```

$\overset{(ii)}{=}$

```
(do (u₀) ← F
    return (u₀ x y))
```

Let us simplify the second part of the diagram. We now use *(i)* the hypothesis, *(ii)* the Frobenius equation and exchange, and *(iii)* the assumption that F has full support.

```
(do (u x y) ← P
    () ← (observe u u0)
    () ← (observe x a)
    return (y))
```

$\overset{(i)}{=}$

```
(do (u) ← F
    (x) ← (G x)
    (y) ← (H x u)
    () ← (observe u u0)
    () ← (observe x a)
    return (y))
```

$\overset{(ii)}{=}$

```
(do
  (u) ← F
  () ← (observe u u0)
  (x) ← (G x)
  (y) ← (H x u0)
  () ← (observe x a)
  return (y))
```

$\overset{(iii)}{=}$

```
(do
  (x) ← (G x)
  (y) ← (H x u0)
  () ← (observe x a)
  return (y))
```

Let us conclude the proof. We *(i)* substitute both simplifications, *(ii)* we use the totality of H, and *(iii)* the Frobenius equation and the totality of G.

```
(do (u0 x0 y0) ← P
    y ← (do
        (u x y) ← P
        () ← (observe u u0)
        () ← (observe x a)
        return (y))
    return (y))
```

$\overset{(i)}{=}$

```
(do (u0) ← F
    y ← (do
        (x) ← (G x)
        () ← (observe x a)
        (y) ← (H x u0)
        return (y))
    return (y))
```

$\overset{(ii)}{=}$

```
(do (u0) ← F
    (x) ← (do
        (x) ← (G x)
        () ← (observe x a)
        return (x))
    (y) ← (H x u0)
    return (y))
```

$\overset{(ii)}{=}$

```
(do (u0) ← F
    (y) ← (H a u0)
    return (y))
```

This concludes the proof: we have shown that the interventional formula can be written in terms of P, exclusively. □

Single-point extensions

Definition .4 (Extension)**.** Given two monads, $(\mathbf{S}, \mu^{\mathbf{S}}, \eta^{\mathbf{S}})$ and $(\mathbf{T}, \mu^{\mathbf{T}}, \eta^{\mathbf{T}})$, an $\mathbf{S}$-*extension* of the $\mathbf{T}$-algebra (X, α) is a $\mathbf{T}$-algebra of the form $(\mathbf{S}X, \alpha^{\#})$ such that $\alpha \fatsemi \eta^{\mathbf{S}} = \mathbf{T}(\eta^{\mathbf{S}}) \fatsemi \alpha^{\#}$.

$$\begin{array}{ccc} \mathbf{T}X & \xrightarrow{\mathbf{T}\eta^{\mathbf{S}}} & \mathbf{TS}X \\ {\scriptstyle\alpha}\downarrow & & \downarrow{\scriptstyle\alpha^{\#}} \\ X & \xrightarrow{\eta^{\mathbf{S}}} & \mathbf{S}X \end{array}$$

Remark .5. A *convex algebra* is an algebra for the finitely-supported distribution monad (D). A *one-point extension* of a convex algebra — in the sense of Sokolova and Woracek [58] — is precisely an M-extension, an extension over the maybe monad.

Theorem .6 (Black-hole is unique [58, Theorem 5.3])**.** *The black-hole one-point extension, defined by*

$$\alpha^{\#}(d) = \begin{cases} d, & \text{if } d(\bot) = 0, \\ \bot, & \text{if } d(\bot) > 0, \end{cases}$$

is the only functorial one-point extension of convex algebras.

It is well-known that a distributive law, $\psi \colon \mathbf{TS} \to \mathbf{ST}$, gives a lifting of the first monad, $\mathbf{S}$, to the category of algebras of second, $\mathbf{T}$ [2]. The lifted monad acts on algebras by extending them. Let us recall this construction here.

Proposition .7 (Extensions from distributive laws)**.** *Any distributive law, $\psi \colon \mathbf{TS} \to \mathbf{ST}$, determines a functorial $\mathbf{S}$-extension of $\mathbf{T}$-algebras.*

Proof. Given a distributive law, $\psi \colon \mathbf{TS} \to \mathbf{ST}$ and a T-algebra (X, α), consider the morphism $\alpha^{\#} = \psi_X \fatsemi \mathbf{S}(\alpha)$. The lifted monad, $\ell(\mathbf{S})$, following Beck's work [2], acts on objects (X, α) by $\ell(\mathbf{S})(X, \alpha) = (\mathbf{S}(X), \alpha^{\#})$. Let us now check that $(\mathbf{S}X, \alpha^{\#})$ is an $\mathbf{S}$-extension of (X, α). We use *(i)* Beck's construction, *(ii)* $\mathbf{S}$-unitality of the distributive law, and *(iii)* naturality.

$$\mathbf{T}(\eta^{\mathbf{S}}) \fatsemi \alpha^{\#} \overset{(i)}{=} \mathbf{T}(\eta^{\mathbf{S}}) \fatsemi \psi \fatsemi \mathbf{S}(\alpha) \overset{(ii)}{=} \eta^{S} \fatsemi \mathbf{S}(\alpha) \overset{(iii)}{=} \alpha \fatsemi \eta^{\mathbf{S}}.$$

This concludes the proof. □