What Monads Can and Cannot Do with a bit of Extra Time

Rasmus Møgelberg & Maaike Zwart

IT University of Copenhagen

21 February 2024

Does the Powerset monad distribute over the Delay monad?

Does the Powerset monad distribute over the Delay monad?

We will:

Try the obvious parallel computation

Does the Powerset monad distribute over the Delay monad?

- Try the obvious parallel computation
- See that it fails Rasmus Møgelberg and Andrea Vezzosi

Does the Powerset monad distribute over the Delay monad?

- Try the obvious parallel computation
- See that it fails Rasmus Møgelberg and Andrea Vezzosi
- Discover why it fails not because of idempotence!

Does the Powerset monad distribute over the Delay monad?

- Try the obvious parallel computation
- See that it fails Rasmus Møgelberg and Andrea Vezzosi
- Discover why it fails not because of idempotence!
- Prove that it is impossible ...because of idempotence.

Why use monads?

Why combine monads?

What are monads?



Why use monads?

Models of Computation: non-determinism, probability, ...
Why combine monads?

What are monads?

Why use monads?

Models of Computation: non-determinism, probability, ...

Why combine monads?

Modelling multiple computational effects.

What are monads?

Why use monads?

Models of Computation: non-determinism, probability, ...

Why combine monads?

Modelling multiple computational effects.

What are monads?

Functors with some structure:

$$\langle \mathcal{M}, \eta : 1 \to \mathcal{M}, \mu : \mathcal{M} \mathcal{M} \to \mathcal{M} \rangle$$

Why use monads?

Models of Computation: non-determinism, probability, ...

Why combine monads?

Modelling multiple computational effects.

What are monads?

Functors with some structure:

$$\langle \mathcal{M}, \eta : 1 \to \mathcal{M}, \mu : \mathcal{M} \mathcal{M} \to \mathcal{M} \rangle$$

Powerset monad for non-determinism:

$$\langle \mathcal{P}, \eta_{\mathcal{P}}, \mu_{\mathcal{P}} \rangle$$

$$\mathcal{P}(X) = \{Y | Y \subseteq X \text{ finite}\}$$

 $\eta_{\mathcal{P}}(x) = \{x\}$
 $\mu_{\mathcal{P}}(Y) = \bigcup Y$

Why use monads?

Models of Computation: non-determinism, probability, ...

Why combine monads?

Modelling multiple computational effects.

What are monads?

Functors with some structure:

$$\langle \mathcal{M}, \eta : 1 \to \mathcal{M}, \mu : \mathcal{M}\mathcal{M} \to \mathcal{M} \rangle$$

 $\{1^{(0)}, *^{(2)}\}$

Powerset monad for non-determinism:

$$\langle \mathcal{P}, \eta_{\mathcal{P}}, \mu_{\mathcal{P}} \rangle$$

$$\mathcal{P}(X) = \{Y | Y \subseteq X \text{ finite}\}$$

$$\eta_{\mathcal{P}}(x) = \{x\}$$

$$\mu_{\mathcal{P}}(Y) = \bigcup Y$$

$$1 * x = x$$

$$x * (y * z) = (x * y) * z$$

$$x * y = y * x$$

$$x * x = x$$

Delay Monad

For recursion / computation steps

Coinductive version:

$$\langle \mathcal{D}, \eta_{\mathcal{D}}, \mu_{\mathcal{D}} \rangle$$

 $\langle \mathcal{D}, \eta_{\mathcal{D}}, \mu_{\mathcal{D}} \rangle$

$$\mathcal{D}(X) \simeq X + \mathcal{D}(X)$$

 $\eta_{\mathcal{D}}(x) = \text{now } x = \text{inl } x$
 $\text{step } x = \text{inr } x$
 $\mu_{\mathcal{D}}(d) = \text{`adding steps'}$

Delay Monad

For recursion / computation steps

Coinductive version:

$$\mathcal{D}(X) \simeq X + \mathcal{D}(X)$$

 $\eta_{\mathcal{D}}(x) = \text{now } x = \text{inl } x$

 $\langle \mathcal{D}, \eta_{\mathcal{D}}, \mu_{\mathcal{D}} \rangle$

$$\mu_{\mathcal{D}}(d) = \text{`adding steps'}$$

step x = inr x

 $\mu_{\mathcal{D}}(\operatorname{step} \operatorname{now}(\operatorname{step} \operatorname{step} \operatorname{now} x)) = \operatorname{step} \operatorname{step} \operatorname{step} \operatorname{now} x$

To make a monad \mathcal{DP} , need distributive law $\lambda : \mathcal{PD} \to \mathcal{DP}$.

Set of computations:

To make a monad \mathcal{DP} , need distributive law $\lambda : \mathcal{PD} \to \mathcal{DP}$.

Set of computations:

To make a monad \mathcal{DP} , need distributive law $\lambda : \mathcal{PD} \to \mathcal{DP}$.

Set of computations:

To make a monad \mathcal{DP} , need distributive law $\lambda : \mathcal{PD} \to \mathcal{DP}$.

Set of computations:

{?,5,?,?,}

To make a monad \mathcal{DP} , need distributive law $\lambda : \mathcal{PD} \to \mathcal{DP}$.

Set of computations:

Computation of a set, via parallel computation:

To make a monad \mathcal{DP} , need distributive law $\lambda : \mathcal{PD} \to \mathcal{DP}$.

Set of computations:

Computation of a set, via parallel computation:

To make a monad \mathcal{DP} , need distributive law $\lambda : \mathcal{PD} \to \mathcal{DP}$.

Set of computations:

Computation of a set, via parallel computation:

To make a monad \mathcal{DP} , need distributive law $\lambda : \mathcal{PD} \to \mathcal{DP}$.

Set of computations:

Computation of a set, via parallel computation:

$$\{2,5,3,8,6\}$$

Total time: max of computation times

More precise:

$$\begin{split} &\lambda\{\mathsf{now}\,x,\mathsf{now}\,y\} = \mathsf{now}\{x,y\} \\ &\lambda\{\mathsf{step}\,d,\mathsf{now}\,y\} = \mathsf{step}(\lambda\{d,\mathsf{now}\,y\}) \\ &\lambda\{\mathsf{step}\,d,\mathsf{step}\,d'\} = \mathsf{step}(\lambda\{d,d'\}) \end{split}$$

so:

 $\{\text{step step now } x, \text{now } y, \text{step now } z\} \mapsto \text{step step now} \{x, y, z\}$

More precise:

$$\begin{split} &\lambda\{\mathsf{now}\,x,\mathsf{now}\,y\} = \mathsf{now}\{x,y\} \\ &\lambda\{\mathsf{step}\,d,\mathsf{now}\,y\} = \mathsf{step}(\lambda\{d,\mathsf{now}\,y\}) \\ &\lambda\{\mathsf{step}\,d,\mathsf{step}\,d'\} = \mathsf{step}(\lambda\{d,d'\}) \end{split}$$

so:

```
\{\mathsf{step}\,\mathsf{step}\,\mathsf{now}\,x,\mathsf{now}\,y,\mathsf{step}\,\mathsf{now}\,z\} \mapsto \mathsf{step}\,\mathsf{step}\,\mathsf{now}\{x,y,z\}
```

Not a distributive law!

More precise:

```
\begin{split} &\lambda\{\mathsf{now}\,x,\mathsf{now}\,y\} = \mathsf{now}\{x,y\} \\ &\lambda\{\mathsf{step}\,d,\mathsf{now}\,y\} = \mathsf{step}(\lambda\{d,\mathsf{now}\,y\}) \\ &\lambda\{\mathsf{step}\,d,\mathsf{step}\,d'\} = \mathsf{step}(\lambda\{d,d'\}) \end{split}
```

so:

```
\{\mathsf{step}\,\mathsf{step}\,\mathsf{now}\,x,\mathsf{now}\,y,\mathsf{step}\,\mathsf{now}\,z\} \mapsto \mathsf{step}\,\mathsf{step}\,\mathsf{now}\{x,y,z\}
```

Not a distributive law! Why?

More precise:

```
\begin{split} &\lambda\{\mathsf{now}\,x,\mathsf{now}\,y\} = \mathsf{now}\{x,y\} \\ &\lambda\{\mathsf{step}\,d,\mathsf{now}\,y\} = \mathsf{step}(\lambda\{d,\mathsf{now}\,y\}) \\ &\lambda\{\mathsf{step}\,d,\mathsf{step}\,d'\} = \mathsf{step}(\lambda\{d,d'\}) \end{split}
```

so:

```
\{\operatorname{step}\operatorname{step}\operatorname{now} x,\operatorname{now} y,\operatorname{step}\operatorname{now} z\} \mapsto \operatorname{step}\operatorname{step}\operatorname{now}\{x,y,z\}
```

Not a distributive law! Why? NOT idempotence!

$$\begin{array}{ccc} \mathcal{PDD} & \xrightarrow{\lambda\mathcal{D}} & \mathcal{DPD} & \xrightarrow{\mathcal{D}\lambda} & \mathcal{DDP} \\ \Big\downarrow^{\mathcal{P}\mu^{\mathcal{D}}} & & & \mu^{\mathcal{D}}\mathcal{P} \Big\downarrow \\ \mathcal{PD} & \xrightarrow{\lambda} & & \mathcal{DP} \end{array}$$

 $\{\text{step now}(\text{now }x), \text{now}(\text{step now }y)\}$

$$\begin{array}{ccc} \mathcal{PDD} & \xrightarrow{\lambda\mathcal{D}} & \mathcal{DPD} & \xrightarrow{\mathcal{D}\lambda} & \mathcal{DDP} \\ & \downarrow^{\mathcal{P}\mu^{\mathcal{D}}} & & \mu^{\mathcal{D}}\mathcal{P} \\ & & \downarrow^{\mathcal{D}}\mathcal{D} & & & & \mathcal{DP} \end{array}$$

 $\{\text{step now}(\text{now }x), \text{now}(\text{step now }y)\} \downarrow_{\mathcal{P}\mu^{\mathcal{D}}} \{\text{step now }x, \text{step now }y\}$

$$\begin{array}{ccc} \mathcal{P}\mathcal{D}\mathcal{D} & \xrightarrow{\lambda\mathcal{D}} & \mathcal{D}\mathcal{P}\mathcal{D} & \xrightarrow{\mathcal{D}\lambda} & \mathcal{D}\mathcal{D}\mathcal{P} \\ \downarrow^{\mathcal{P}\mu^{\mathcal{D}}} & & \mu^{\mathcal{D}}\mathcal{P} \\ \downarrow^{\mathcal{P}}\mathcal{D} & \xrightarrow{\lambda} & \mathcal{D}\mathcal{P} \end{array}$$

$$\{\operatorname{step \, now}(\operatorname{now} x), \operatorname{now}(\operatorname{step \, now} y)\} \downarrow_{\mathcal{P}\mu^{\mathcal{D}}} \{\operatorname{step \, now} x, \operatorname{step \, now} y\}$$

$$\xrightarrow{\lambda} \operatorname{step \, now}\{x,y\}$$

$$\begin{array}{ccc} \mathcal{P}\mathcal{D}\mathcal{D} & \xrightarrow{\lambda\mathcal{D}} & \mathcal{D}\mathcal{P}\mathcal{D} & \xrightarrow{\mathcal{D}\lambda} & \mathcal{D}\mathcal{D}\mathcal{P} \\ \downarrow^{\mathcal{P}\mu^{\mathcal{D}}} & & \mu^{\mathcal{D}}\mathcal{P} \\ \downarrow^{\mathcal{P}}\mathcal{D} & \xrightarrow{\lambda} & \mathcal{D}\mathcal{P} \end{array}$$

$$\{ \mathsf{step} \, \mathsf{now}(\mathsf{now} \, x), \, \mathsf{now}(\mathsf{step} \, \mathsf{now} \, y) \} \, \downarrow_{\mathcal{P}\mu^{\mathcal{D}}} \{ \mathsf{step} \, \mathsf{now} \, x, \mathsf{step} \, \mathsf{now} \, y \}$$

$$\qquad \qquad \qquad \xrightarrow{\lambda} \mathsf{step} \, \mathsf{now} \{ x, y \}$$

 $\{\text{step now}(\text{now }x), \text{now}(\text{step now }y)\}$

$$\begin{array}{ccc} \mathcal{P}\mathcal{D}\mathcal{D} & \xrightarrow{\lambda\mathcal{D}} & \mathcal{D}\mathcal{P}\mathcal{D} & \xrightarrow{\mathcal{D}\lambda} & \mathcal{D}\mathcal{D}\mathcal{P} \\ \downarrow^{\mathcal{P}\mu^{\mathcal{D}}} & & \mu^{\mathcal{D}}\mathcal{P} \\ \downarrow^{\mathcal{P}}\mathcal{D} & \xrightarrow{\lambda} & \mathcal{D}\mathcal{P} \end{array}$$

$$\{ \mathsf{step} \, \mathsf{now}(\mathsf{now} \, x), \, \mathsf{now}(\mathsf{step} \, \mathsf{now} \, y) \} \, \downarrow_{\mathcal{P}\mu^{\mathcal{D}}} \{ \mathsf{step} \, \mathsf{now} \, x, \mathsf{step} \, \mathsf{now} \, y \}$$

$$\qquad \qquad \qquad \xrightarrow{\lambda} \mathsf{step} \, \mathsf{now} \{ x, y \}$$

 $\{\text{step now}(\text{now }x), \text{now}(\text{step now }y)\} \xrightarrow{\lambda \mathcal{D}} \text{step now}(\{\text{now }x, \text{step now }y\})$

$$\begin{array}{ccc} \mathcal{P}\mathcal{D}\mathcal{D} & \xrightarrow{\lambda\mathcal{D}} & \mathcal{D}\mathcal{P}\mathcal{D} & \xrightarrow{\mathcal{D}\lambda} & \mathcal{D}\mathcal{D}\mathcal{P} \\ \downarrow^{\mathcal{P}\mu^{\mathcal{D}}} & & \mu^{\mathcal{D}}\mathcal{P} \\ \downarrow^{\mathcal{P}}\mathcal{D} & \xrightarrow{\lambda} & \mathcal{D}\mathcal{P} \end{array}$$

$$\{ \mathsf{step} \, \mathsf{now}(\mathsf{now} \, x), \, \mathsf{now}(\mathsf{step} \, \mathsf{now} \, y) \} \, \downarrow_{\mathcal{P}\mu^{\mathcal{D}}} \{ \mathsf{step} \, \mathsf{now} \, x, \mathsf{step} \, \mathsf{now} \, y \}$$

$$\qquad \qquad \qquad \xrightarrow{\lambda} \mathsf{step} \, \mathsf{now} \{ x, y \}$$

$$\{\operatorname{step \ now}(\operatorname{now} x), \operatorname{now}(\operatorname{step \ now} y)\} \xrightarrow{\lambda \mathcal{D}} \operatorname{step \ now}(\{\operatorname{now} x, \operatorname{step \ now} y\})$$

$$\xrightarrow{\mathcal{D}\lambda} \operatorname{step \ now}(\operatorname{step \ now}\{x,y\})$$

Parallel Computation

$$\begin{array}{ccc} \mathcal{P}\mathcal{D}\mathcal{D} & \xrightarrow{\lambda\mathcal{D}} & \mathcal{D}\mathcal{P}\mathcal{D} & \xrightarrow{\mathcal{D}\lambda} & \mathcal{D}\mathcal{D}\mathcal{P} \\ \downarrow^{\mathcal{P}\mu^{\mathcal{D}}} & & \mu^{\mathcal{D}}\mathcal{P} \\ \downarrow^{\mathcal{P}}\mathcal{D} & \xrightarrow{\lambda} & \mathcal{D}\mathcal{P} \end{array}$$

```
 \{ \mathsf{step} \, \mathsf{now}(\mathsf{now} \, x), \, \mathsf{now}(\mathsf{step} \, \mathsf{now} \, y) \} \, \downarrow_{\mathcal{P}\mu^{\mathcal{D}}} \{ \mathsf{step} \, \mathsf{now} \, x, \mathsf{step} \, \mathsf{now} \, y \}   \qquad \qquad \qquad \xrightarrow{\lambda} \mathsf{step} \, \mathsf{now} \{ x, y \}
```

```
 \begin{split} \{ \operatorname{step \, now}(\operatorname{now} x), \operatorname{now}(\operatorname{step \, now} y) \} & \xrightarrow{\lambda \mathcal{D}} \operatorname{step \, now}(\{\operatorname{now} x, \operatorname{step \, now} y\}) \\ & \xrightarrow{\mathcal{D} \lambda} \operatorname{step \, now}(\operatorname{step \, now}\{x, y\}) \\ & \downarrow_{\mu^{\mathcal{D}}\mathcal{P}} \operatorname{step \, step \, now}\{x, y\} \end{split}
```

Parallel Computation

So what went wrong? Nothing specific to Powerset! Only ingredient: "structure with two elements".

Theorem

Parallel computation is **never** a distributive law for $\mathcal{MD} \to \mathcal{DM}$ if \mathcal{M} is presented by a theory with a binary term.

```
\begin{split} &\lambda\{\mathsf{now}\,x,\mathsf{now}\,y\} = \mathsf{now}\{x,y\} \\ &\lambda\{\mathsf{step}\,d,\mathsf{now}\,y\} = \mathsf{step}(\lambda\{d,\mathsf{now}\,y\}) \\ &\lambda\{\mathsf{step}\,d,\mathsf{step}\,d'\} = \mathsf{step}(\lambda\{d,d'\}) \end{split}
```

```
\lambda \{ \mathsf{now}\, x, \mathsf{now}\, y \} = \mathsf{now} \{ x, y \}
\lambda \{ \mathsf{step}\, d, \mathsf{now}\, y \} = \mathsf{step} (\lambda \{ d, \mathsf{now}\, y \})
\underline{\lambda \{ \mathsf{step}\, d, \mathsf{step}\, d' \}} = \mathsf{step} (\lambda \{ d, d' \})
```

$$\lambda \{\mathsf{now}\,x,\mathsf{now}\,y\} = \mathsf{now}\{x,y\}$$
$$\lambda \{\mathsf{step}\,d,\mathsf{now}\,y\} = \mathsf{step}(\lambda \{d,\mathsf{now}\,y\})$$
$$\lambda \{\mathsf{step}\,d,\mathsf{step}\,d'\} = \mathsf{step}(\lambda \{d,d'\})$$

$$\lambda \{\mathsf{now}\,x,\mathsf{now}\,y\} = \mathsf{now}\{x,y\}$$
$$\lambda \{\mathsf{step}\,d,\mathsf{now}\,y\} = \mathsf{step}(\lambda \{d,\mathsf{now}\,y\})$$
$$\lambda \{\mathsf{step}\,d,\mathsf{step}\,d'\} = \mathsf{step}(\lambda \{d,d'\})$$

$$\lambda\{\mathsf{now}\,x,\mathsf{now}\,y\} = \mathsf{now}\{x,y\}$$
$$\lambda\{\mathsf{step}\,d,\mathsf{now}\,y\} = \mathsf{step}(\lambda\{d,\mathsf{now}\,y\})$$
$$\lambda\{\mathsf{step}\,d,\mathsf{step}\,d'\} = \mathsf{step}(\lambda\{d,d'\})$$

$$\begin{split} &\lambda\{\mathsf{now}\,x,\mathsf{now}\,y\} = \mathsf{now}\{x,y\} \\ &\lambda\{\mathsf{step}\,d,\mathsf{now}\,y\} = \mathsf{step}(\lambda\{d,\mathsf{now}\,y\}) \\ &\lambda\{\mathsf{step}\,d,\mathsf{step}\,d'\} = \mathsf{step}(\lambda\{d,d'\}) \end{split}$$

$$\begin{split} &\lambda\{\mathsf{now}\,x,\mathsf{now}\,y\} = \mathsf{now}\{x,y\} \\ &\lambda\{\mathsf{step}\,d,\mathsf{now}\,y\} = \mathsf{step}(\lambda\{d,\mathsf{now}\,y\}) \\ &\lambda\{\mathsf{step}\,d,\mathsf{step}\,d'\} = \mathsf{step}(\lambda\{d,d'\}) \end{split}$$

$$\begin{split} &\lambda\{\mathsf{now}\,x,\mathsf{now}\,y\} = \mathsf{now}\{x,y\} \\ &\lambda\{\mathsf{step}\,d,\mathsf{now}\,y\} = \mathsf{step}(\lambda\{d,\mathsf{now}\,y\}) \\ &\lambda\{\mathsf{step}\,d,\mathsf{step}\,d'\} = \mathsf{step}(\lambda\{d,d'\}) \end{split}$$

No Hope in General?

Theorem

There is no causal distributive law $\mathcal{MD} \to \mathcal{DM}$, if \mathcal{M} is presented by a theory with an idempotent and commutative term.

No Hope in General?

Theorem

There is no causal distributive law $\mathcal{MD} \to \mathcal{DM}$, if \mathcal{M} is presented by a theory with an idempotent and commutative term.

$$\mathcal{M}\mathcal{D}^{\kappa} \to \mathcal{D}^{\kappa} \mathcal{M} \Rightarrow \mathcal{M}\mathcal{D} \to \mathcal{D}\mathcal{M}$$

No Hope in General?

Theorem

There is no causal distributive law $\mathcal{MD} \to \mathcal{DM}$, if \mathcal{M} is presented by a theory with an idempotent and commutative term.

$$\mathcal{M}\mathcal{D}^{\kappa} \to \mathcal{D}^{\kappa}\mathcal{M} \Rightarrow \mathcal{M}\mathcal{D} \to \mathcal{D}\mathcal{M}$$
$$\mathcal{M}\mathcal{D} \to \mathcal{D}\mathcal{M} \Rightarrow \mathcal{M}\mathcal{D}^{\kappa} \to \mathcal{D}^{\kappa}\mathcal{M}$$

Example in the paper.

Conclusion

We saw:

- Parallel computation fails to give a distributive law.
- Distributive law $\mathcal{PD} \to \mathcal{DP}$ impossible.
- Idempotence destroys causality.

Conclusion

We saw:

- Parallel computation fails to give a distributive law.
- Distributive law $\mathcal{PD} \to \mathcal{DP}$ impossible.
- Idempotence destroys causality.

More in the paper!

- Sequential computation:
 - Fails because of idempotence.
 - Works for balanced equations.
- Dist laws for Exceptions, Reader, State, Selection.

Conclusion

We saw:

- Parallel computation fails to give a distributive law.
- Distributive law $\mathcal{PD} \to \mathcal{DP}$ impossible.
- Idempotence destroys causality.

More in the paper!

- Sequential computation:
 - Fails because of idempotence.
 - Works for balanced equations.
- Dist laws for Exceptions, Reader, State, Selection.

A bit of hope:

Parallel computation does give a distributive law up to weak bisimilarity.

