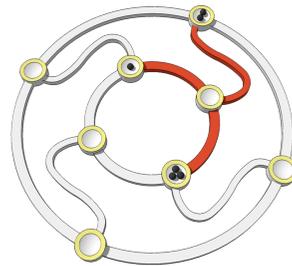


Nondeterminism and probability:  
A tour of probabilistic programming,  
from abstraction to the “refinement  
paradox”.



Annabelle McIver,  
Macquarie University,  
Sydney

## Quick summary

- A short history of probabilistic programming;
- How to build the semantics you want in four easy stages;
- A cute application;
- The refinement paradox, and how probability can help shed light.

# A brief History of Computing

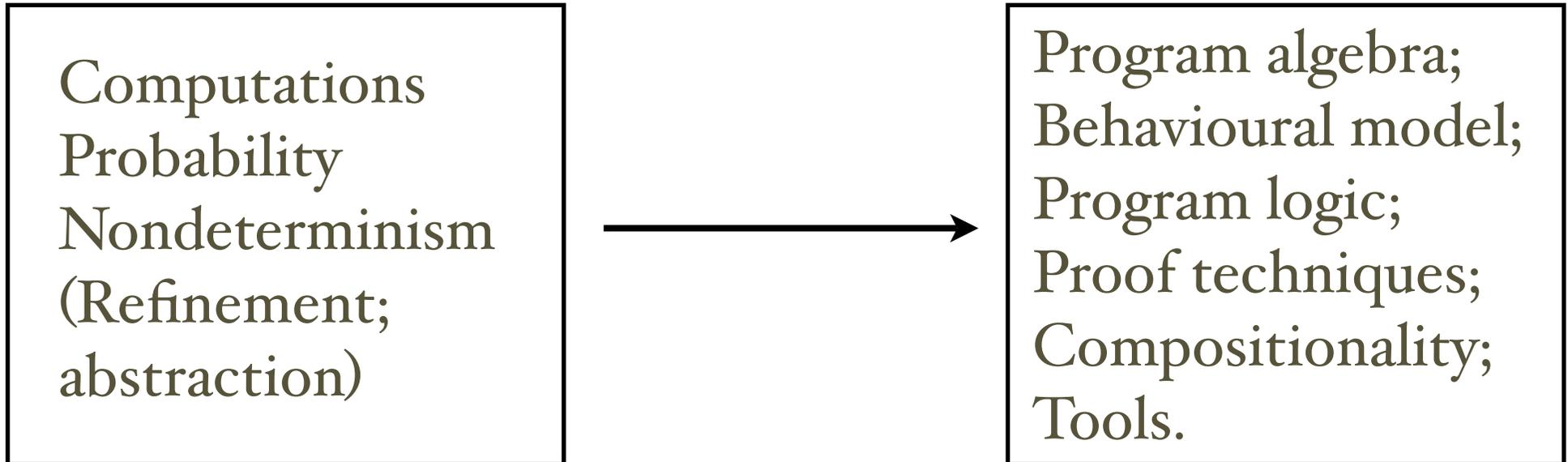
# A brief History of Computing



From working-out the theory...  
to working, in theory.



What we want.



But what does this all mean?  
How do these things interact?

In particular what does it mean for refinement of probability: intuitively you might think that probability should increase up the refinement order...

Good  $1/3 \oplus$  Bad  $\sqsubseteq$  Good  $1/2 \oplus$  Bad

In particular what does it mean for refinement of probability: intuitively you might think that probability should increase up the refinement order...

Good  $1/3 \oplus$

Bad

$\sqsubseteq$

Good  $1/2 \oplus$

Bad



$1/3 \oplus$



$1/2 \oplus$



We're going to look for inspiration from a general mathematical construction called powerdomains.

... but now if we reinstantiate “Good” and “Bad”, shouldn’t we also have this?

Good  $2/3 \oplus$  Bad  $\sqsubseteq$  Good  $1/2 \oplus$  Bad



$2/3 \oplus$



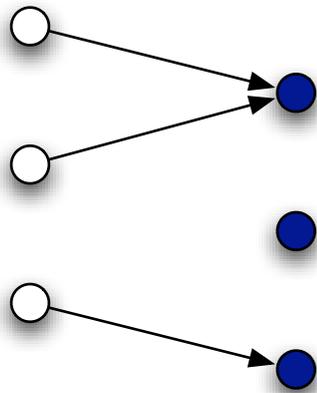
$1/2 \oplus$



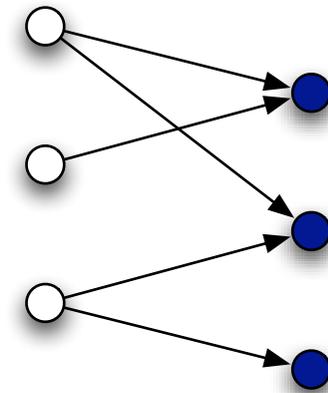
We’re going to look for inspiration from a general mathematical construction called powerdomains to guide the principles underlying refinement.

# Powerdomains

(Originally) a general technique by which a semantic model can be augmented to include nondeterminism in such a way that the underlying computational structure of the original model is maintained.



Functions,  
 $S \longrightarrow S$

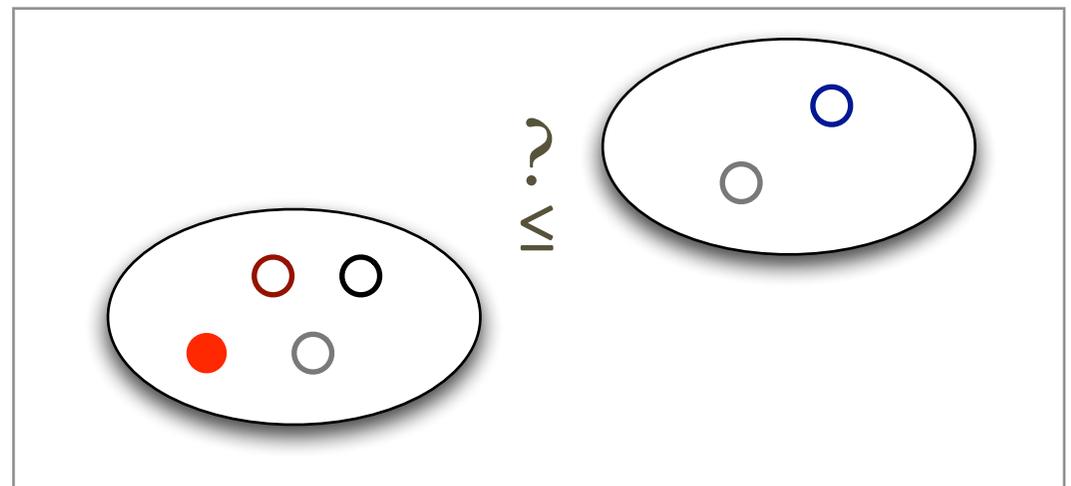
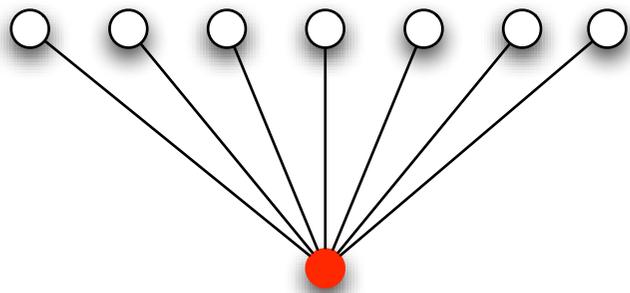


Relations,  
 $S \longrightarrow PS$

# Powerdomains

When we want to distinguish nontermination from other behaviour, we introduce a special “bottom state” ●

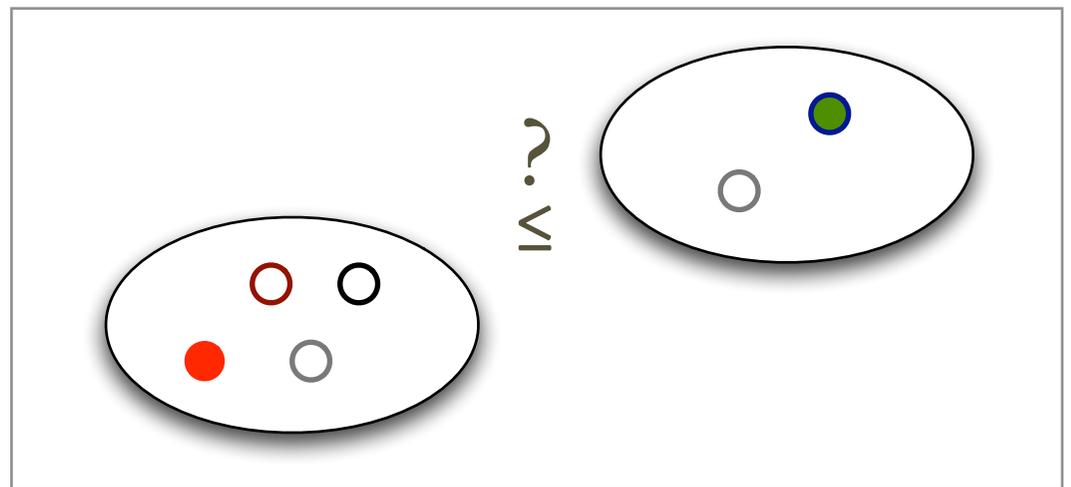
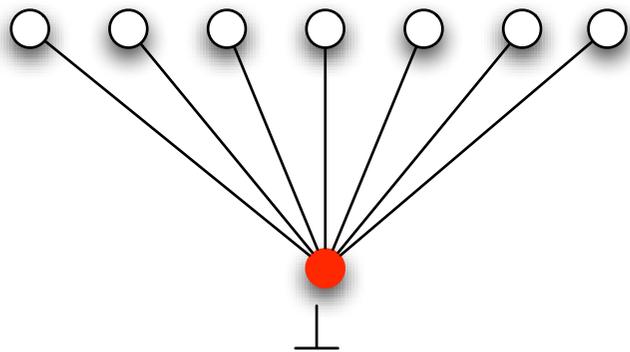
How do we order the programs so that ● is worse than everything, and reducing the range of behaviours corresponds to “more refined”.



# Powerdomains: the Smyth order

$$A \leq_S B \quad \text{iff} \quad (\forall b \in B)(\exists a \in A \cdot a \leq b)$$

$$A \leq_S B \quad \text{iff} \quad (\perp \in A) \vee (B \subseteq A)$$

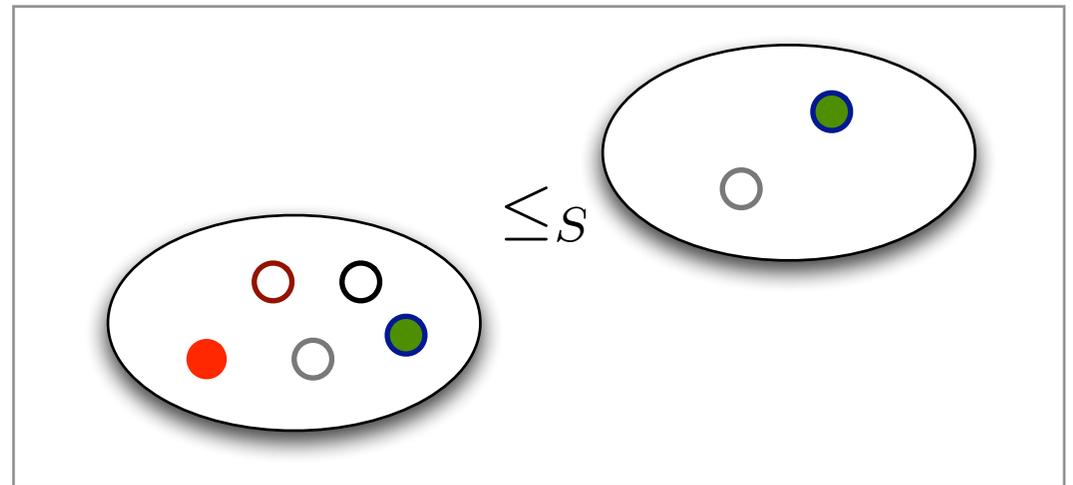


# Powerdomains: the Smyth order

$$A \uparrow \hat{=} \{s \in S_{\perp} \mid (\exists a \in A \cdot a \leq s)\}$$

$\leq_S$  becomes an order (rather than a pre-order) on up-closed sets.

On up-closed sets,  
refinement is simply  
reverse subset  
inclusion.

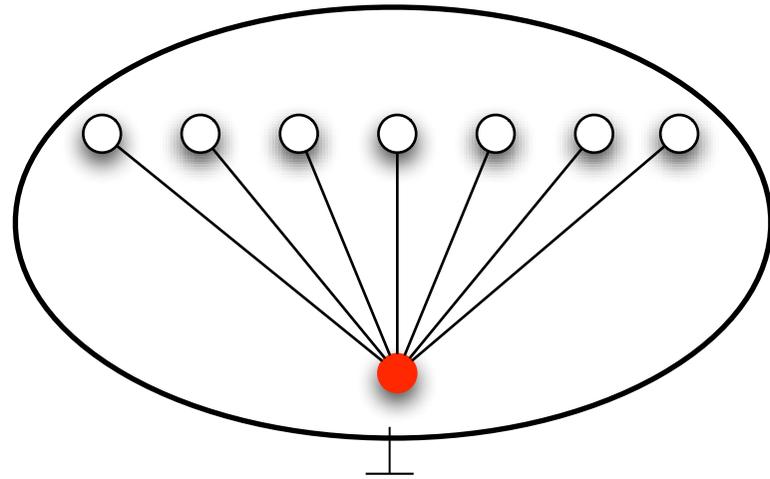
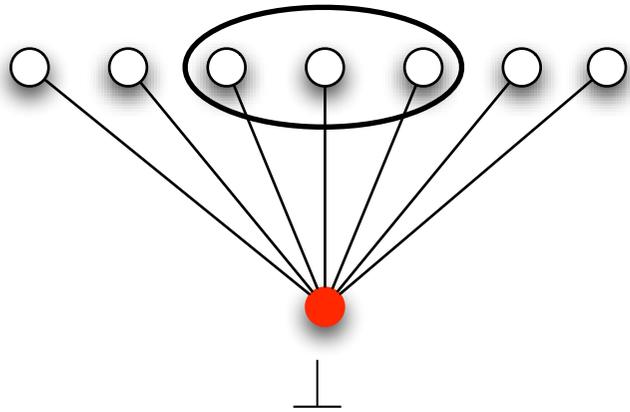


## Probabilistic powerdomains

Given a structure  $(D, \leq)$ , we can construct a powerdomain  $(\text{Eval}.D, \leq)$  where objects are evaluations over  $D$ , and the order is defined to make “appropriate distinctions”.

- Evaluations are real-valued functions which are defined over the open sets of a (fixed) topology; under certain conditions they can be extended to probability distributions.
- Computational domains can be reformulated in terms of the Scott Topology: a set is Scott open if it is “up-closed” and “inaccessible” (any limit of a chain inside the set can only happen if the chain intersects the set).

# Probabilistic powerdomains: Evaluations



$$Eval.S_{\perp} \hat{=} OS_{\perp} \rightarrow [0, 1]$$

Monotone; additive

$$d \leq d' \quad \text{iff} \quad (\forall O \in OS_{\perp} \cdot d.O \leq d'.O)$$

$$d \leq d' \quad \text{iff} \quad (\forall s \in S \cdot d.\{s\} \leq d'.\{s\})$$

Probability can  
increase up the  
refinement order!

# Probabilistic powerdomains: Semantics

Given a structure  $(D, \leq)$ , we can construct a powerdomain  $(\text{Eval}.D, \leq)$  where objects are evaluations over  $D$ , and the order is defined to make “appropriate distinctions”.

Programs

$$S_{\perp} \rightarrow \text{Eval}.S_{\perp}$$

Probabilistic choice

$$(P \oplus_p Q).s \hat{=} p \times P.s + (1-p) \times Q.s$$

Sequence

$$P; Q \hat{=} P \circ Q^{\dagger}$$

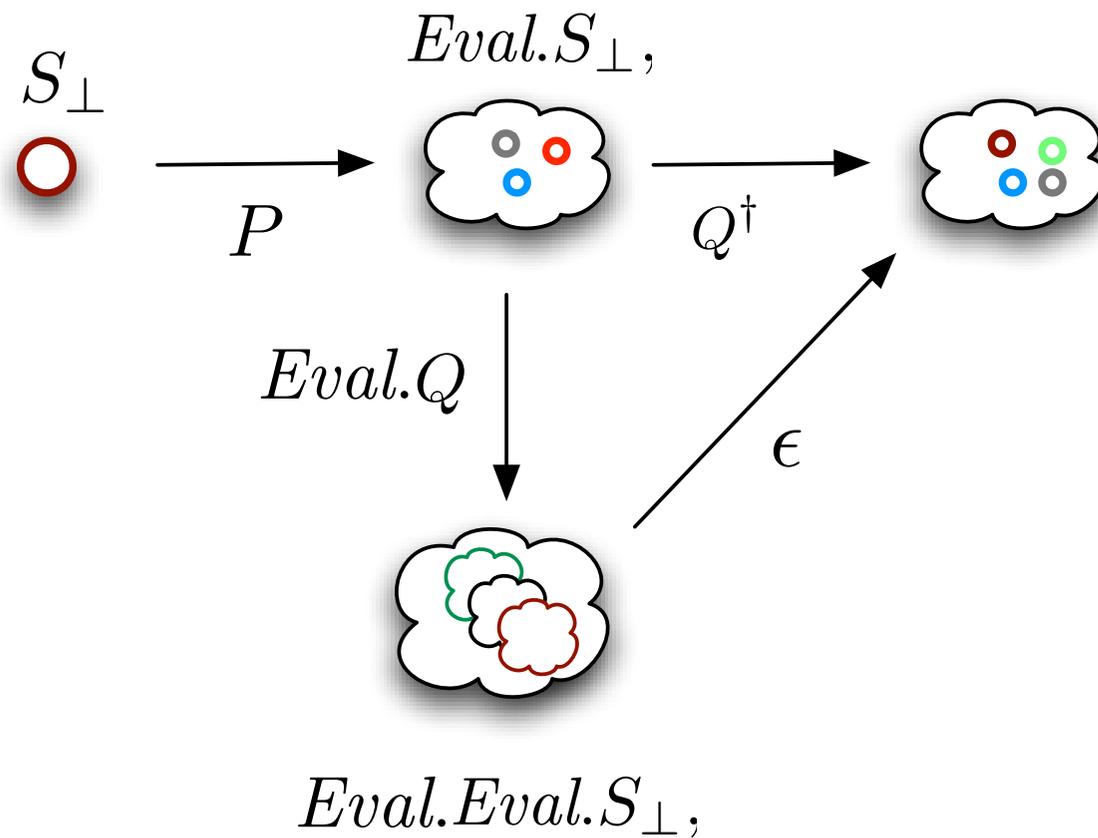
$$Q^{\dagger} : \text{Eval}.S_{\perp} \rightarrow \text{Eval}.S_{\perp}$$

$$Q^{\dagger}.d \hat{=} \sum_{s:S} (d.s) \times Q.s$$

# Probabilistic powerdomains: Defining $Q^\dagger$

Sequence

$$P; Q \hat{=} P \circ Q^\dagger$$



Now we have all the ingredients for instant probabilistic semantics.



“It’s marvelous! You just add water.”

First try:

You will need a flat domain, the Smyth Powerdomain, and the probabilistic powerdomain.

- First add nondeterminism to  $(S_{\perp} \rightarrow S_{\perp}, \sqsubseteq)$
- Next add nondeterminism to get  $(S_{\perp} \rightarrow \mathbb{P}S_{\perp}, \sqsubseteq_S)$
- Finally add probability to get  $(Eval.(S_{\perp} \rightarrow \mathbb{P}S_{\perp}), \sqsubseteq_S)$

First try: What do we get?

• Probabilistic arithmetic  $(P \oplus_p Q) \oplus_q R = P \oplus_{pq} (Q \oplus_{\frac{(1-p)q}{1-pq}} R)$

• Universal probabilistic distributivity....  $(P \oplus_p Q) \sqcap R = (P \sqcap R) \oplus_p (Q \sqcap R)$

.... which implies this ....

$$\begin{aligned}
 & P \sqcap P \\
 = & (a \oplus_{1/2} b) \sqcap (a \oplus_{1/2} b) \\
 = & (a \sqcap (a \oplus_{1/2} b)) \oplus_{1/2} (b \sqcap (a \oplus_{1/2} b)) \\
 = & (a \sqcap a) \oplus_{1/4} ((b \sqcap b) \oplus_{1/3} (a \sqcap b)) \\
 = & a \oplus_{1/4} (b \oplus_{1/3} (a \sqcap b)) \\
 \neq & P
 \end{aligned}$$

$$\begin{aligned}
 & P \hat{=} a \oplus_{1/2} b \\
 & \text{Distribution} \\
 & \text{Distribution, arithmetic} \\
 & a \sqcap a = a \\
 & P \hat{=} a \oplus_{1/2} b
 \end{aligned}$$

!!

# Probability versus nondeterminism

$$(y := 0 \sqcap y := 1); (x := 0 \oplus_{1/2} x := 1)$$

What's the chance that the demon can guess the value of  $x$ ?

# Probability versus nondeterminism

$(y := 0 \sqcap y := 1); (x := 0 \text{ }_{1/2} \oplus x := 1)$  Prob distributes over nondet

=  $(y := 0 \sqcap y := 1); x := 0 \text{ }_{1/2} \oplus (y := 0 \sqcap y := 1); x := 1$

In this model, we can reproduce the demon's choice within each probabilistic branch....

... effectively making the demon able to see into the future.

Whoops!



Next try:

You will need a flat domain, the Smyth Powerdomain, the probabilistic powerdomain, and compactness and convexity.

- First add probability to get  $(Eval.S_{\perp}, \leq)$
- Next add nondeterminism to get  $(S_{\perp} \rightarrow \mathbb{P}Eval.S_{\perp}, \sqsubseteq_P)$
- We need some extra closure conditions:
  - (a) up-closed - for termination
  - (b) Convex closed -  $P_p \oplus P = P$
  - (c) Compact - so that iteration can be approximated by “finite” computations.

As before, refinement is reverse subset inclusion

# Relational-style semantics for a small sequential language

<i>identity</i>	$\llbracket \text{skip} \rrbracket .s$	$\hat{=} \{\bar{s}\}$
<i>assignment</i>	$\llbracket x := a \rrbracket .s$	$\hat{=} \overline{\{s[x \mapsto a]\}}$
<i>composition</i>	$\llbracket P; P' \rrbracket .s$	$\hat{=} \{\sum_{s': S} d.s' \times f'.s' \mid d \in \llbracket P \rrbracket .s; f' \sqsubseteq \llbracket P' \rrbracket\}$ where $f' \in S \rightarrow \bar{S}_\perp$ and in general $f' \sqsubseteq r'$ means $f'.s \in r'.s$ for all $s$ .
<i>choice</i>	$\llbracket \text{if } B \text{ then } P \text{ else } P' \rrbracket .s$	$\hat{=} \text{if } B.s \text{ then } \llbracket P \rrbracket .s \text{ else } \llbracket P' \rrbracket .s$
<i>probability</i>	$\llbracket P \oplus_p P' \rrbracket .s$	$\hat{=} \{d \oplus_p d' \mid d \in \llbracket P \rrbracket .s; d' \in \llbracket P' \rrbracket .s\}$
<i>nondeterminism</i>	$\llbracket P \sqcap P' \rrbracket .s$	$\hat{=} \lceil \llbracket P \rrbracket .s \cup \llbracket P' \rrbracket .s \rceil ,$ where in general $\lceil D \rceil$ is the up-, convex- and Cauchy closure of $D$ .
<i>iteration</i>	$\text{do } G \rightarrow P \text{ od}$	$\hat{=} (\mu X \cdot \text{if } G \text{ then } \llbracket P \rrbracket ; X \text{ else } \llbracket \text{skip} \rrbracket) .$

Some nice laws....

$$P \sqcap P = P \qquad (P \sqcap Q) \oplus_p (P \sqcap R) \sqsubseteq_P P \sqcap (Q \oplus_p R)$$

$$P \sqcap P \sqsubseteq_P P \oplus_p P = P \qquad P \oplus_p (Q \sqcap R) = (P \oplus_p Q) \sqcap (P \oplus_p R)$$

$$P; (Q \oplus_p R) \sqsubseteq_P P; Q \oplus_p P; R$$

$$(Q \oplus_p R); P = (Q; P \oplus_p R; P)$$

This nondeterminism can see what happened after a coin flip, but not before.

# Probability versus nondeterminism

$$(y := 0 \sqcap y := 1); (x := 0 \oplus_{1/2} x := 1)$$

What's the chance that the demon can guess the value of  $x$ ?

# Probability versus nondeterminism

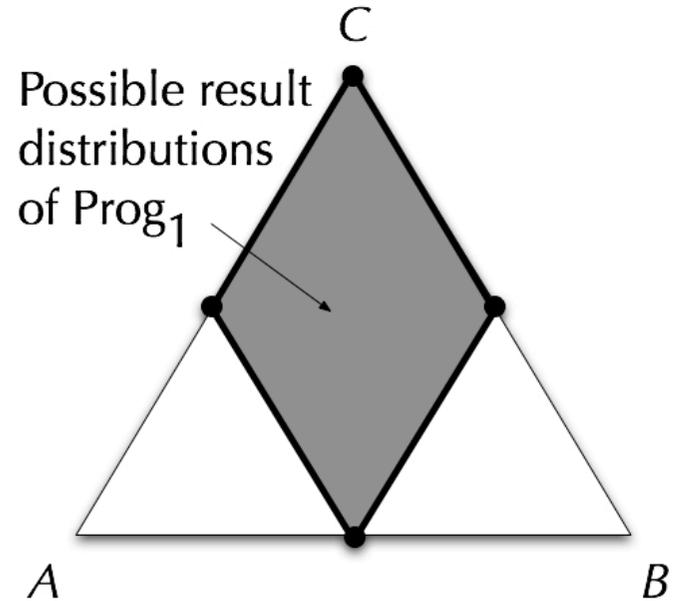
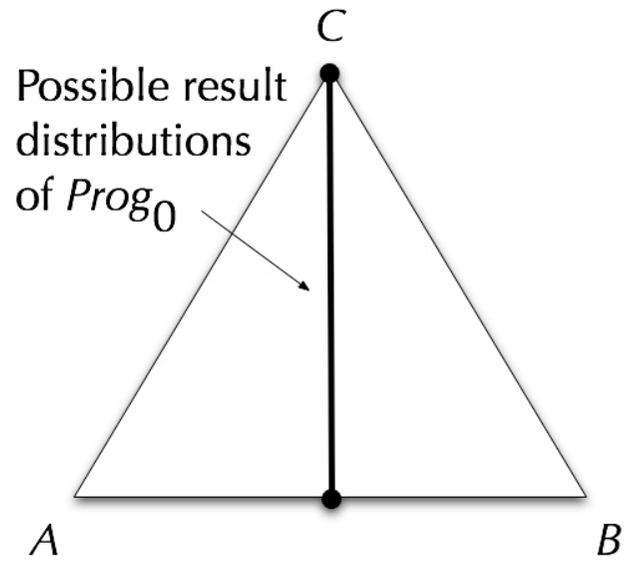
$(y := 0 \sqcap y := 1); (x := 0 \oplus_{1/2} x := 1)$  Nondet distributes over proc

=  $(y := 0); (x := 0 \oplus_{1/2} x := 1) \sqcap (y := 1); (x := 0 \oplus_{1/2} x := 1)$

What's the chance that the demon can guess the value of  $x$ ?

Answer is 1/2.

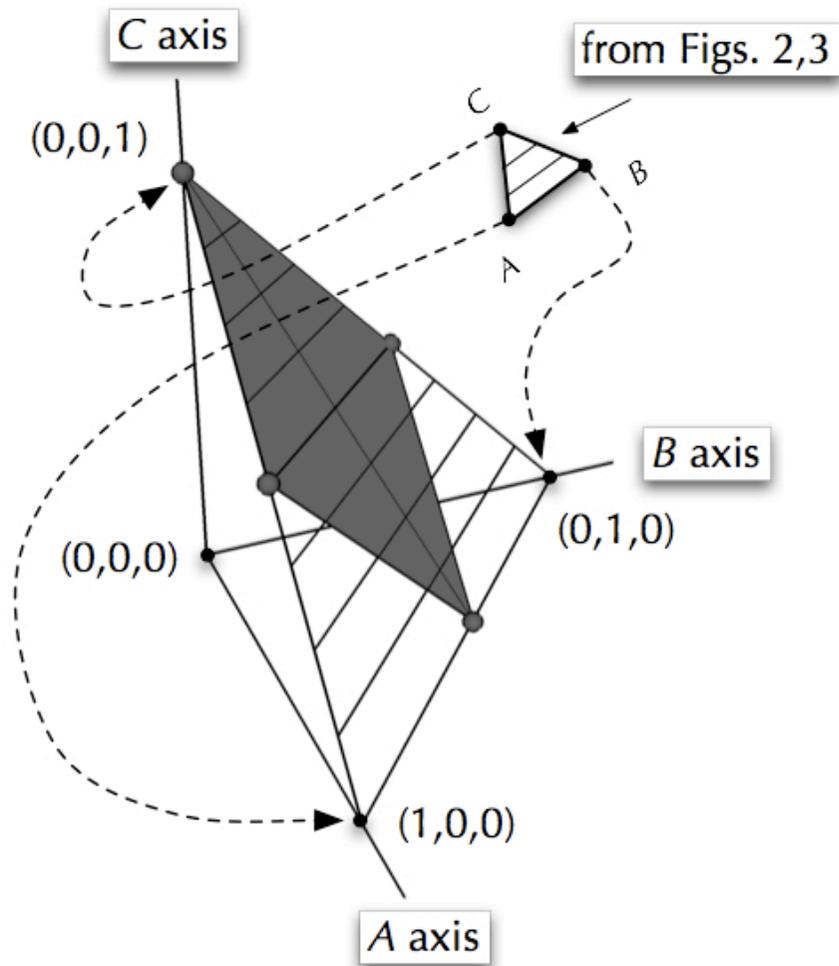
# Geometrical interpretation.



$$Prog_0 \hat{=} (s := A \oplus_{0.5} s := B) \sqcap s := C$$

$$Prog_1 \hat{=} (s := A \sqcap s := C) \oplus_{0.5} (s := B \sqcap s := C)$$

# Geometrical interpretation.



Plotted on the same diagram, we can see immediately the relationship between the two programs.

$$Prog_0 \hat{=} (s := A \oplus_{0.5} s := B) \sqcap s := C$$

$$Prog_1 \hat{=} (s := A \sqcap s := C) \oplus_{0.5} (s := B \sqcap s := C)$$

## Logic and properties

Properties are now  
quantitative

$$\begin{aligned}\mathbb{E}S &\hat{=} S \rightarrow [0, 1] \\ e \leq e' &= (\forall s : S \cdot e.s \leq e'.s)\end{aligned}$$

$$\text{wp}.P.e.s \hat{=} \left( \prod_{d \in P.s} \int_d e \right)$$

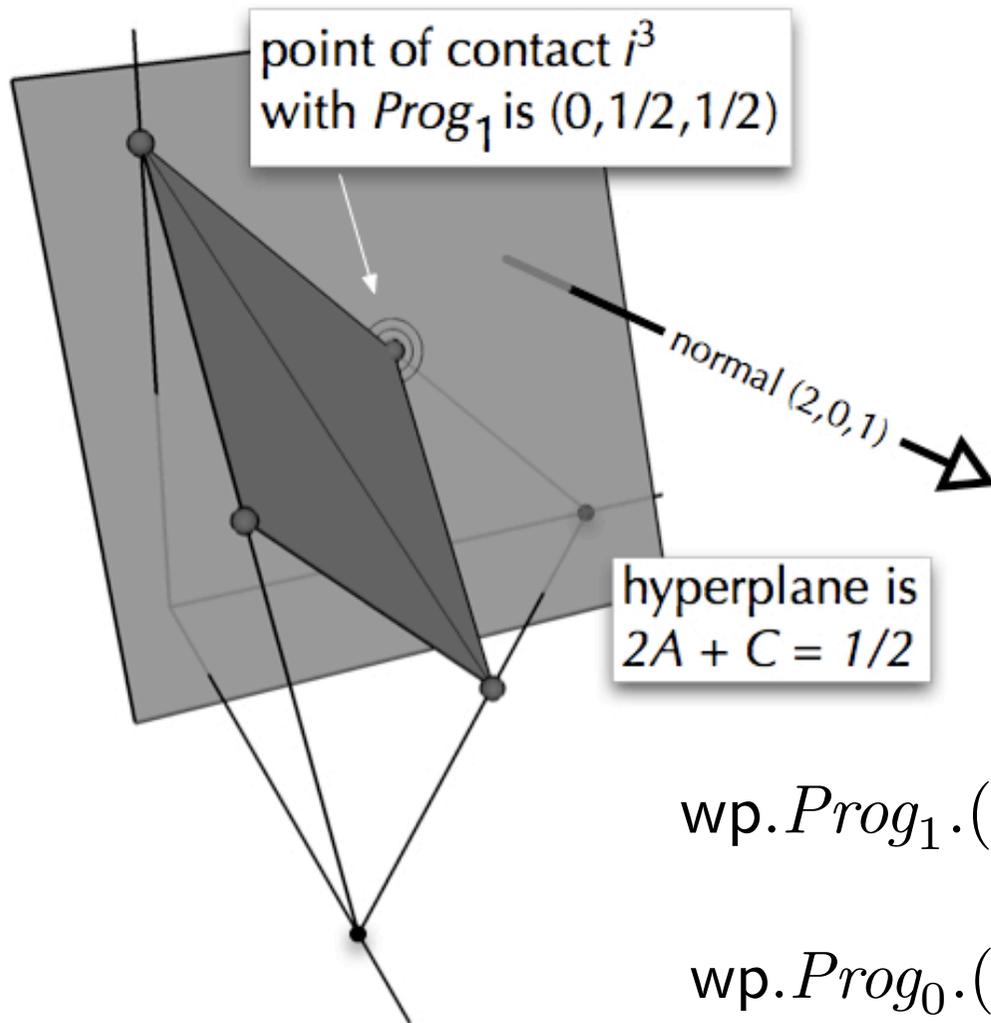
Greatest guaranteed expected value of  $e$  with respect to the results of  $P$  from initial state  $s$ .

$$d \in \text{Eval}S_{\perp}, e \in \mathbb{E}S, \quad \int_d e \hat{=} \sum_{s:S} d.s \times e.s$$

# Transformer semantics for a small sequential language

<i>identity</i>	$\text{wp.skip.expt}$	$\hat{=}$	$\text{expt}$
<i>assignment</i>	$\text{wp.}(x := E).\text{expt}$	$\hat{=}$	$\text{expt}[x := E]$
<i>composition</i>	$\text{wp.}(P; P').\text{expt}$	$\hat{=}$	$\text{wp.P.}(\text{wp.P'.expt})$
<i>choice</i>	$\text{wp.}(\text{if } B \text{ then } P \text{ else } P' \text{ fi}).\text{expt}$		
	$\hat{=}$	$[B] \times \text{wp.P.expt} + [\neg B] \times \text{wp.P'.expt}$	
<i>probability</i>	$\text{wp.}(P_p \oplus P').\text{expt}$		
	$\hat{=}$	$p \times \text{wp.P.expt} + (1-p) \times \text{wp.P'.expt}$	
<i>nondeterminism</i>	$\text{wp.}(P \sqcap P').\text{expt}$	$\hat{=}$	$\text{wp.P.expt} \mathbf{min} \text{wp.P'.expt}$
<i>iteration</i>	$\text{wp.}(\text{do } B \rightarrow r \text{ od}).e \hat{=}$	$(\mu X \cdot [B] \times \text{wp.r.X} + [\neg B] \times e) .$	

# Geometrical interpretation:



Expectations are  
“hyperplanes”.

$$wp.Prog_1 \cdot (2[s = A] + [s = C]) = 1/2$$

$$wp.Prog_0 \cdot (2[s = A] + [s = C]) = 1$$

Logic and properties:  
the monotonic transformers

$$\mathbb{T}S \hat{=} \mathbb{E}S \leftarrow \mathbb{E}S$$

$$[S_{\perp} \rightarrow \mathbb{P}Eval.S_{\perp} \begin{array}{c} \xleftarrow{wp} \\ \xrightarrow{rp} \end{array} \mathbb{T}S$$

$$wp \circ rp = id$$

$$rp \circ wp = id, \text{ if } \left\{ \begin{array}{l} t.(e_p \oplus e') \geq t.e_p \oplus t.e' \\ t.(ke) = kt.e \\ t.(e - k) \geq t.e - k \end{array} \right. \quad \text{“Sublinear”}$$

Why so complicated: can't we just have a whole logic based on probabilities, rather than random variables?

It's a question of compositionality:

$$Prog_0 \hat{=} (s := A \oplus_{0.5} s := B) \sqcap s := C$$

$$Prog_1 \hat{=} (s := A \sqcap s := C) \oplus_{0.5} (s := B \sqcap s := C)$$

Allowed final value(s) of $s$	$A$	$B$	$C$	$A, B$	$B, C$	$C, A$
Maximum possible probability	1/2	1/2	1	1	1	1
Minimum possible probability	0	0	0	0	1/2	1/2

A quantitative logic based on probabilities *is not compositional*.

Consider the following “context”:

$Prog_0; \quad \text{if } s=C \text{ then } (s := A \oplus_{0.5} s := B) \text{ fi}$   
 $Prog_1; \quad \text{if } s=C \text{ then } (s := A \oplus_{0.5} s := B) \text{ fi}$

What’s the probability that the state is A finally?

As we have seen, the two programs can be distinguished in the transformer semantics (by a random variable encoded as an expectation).

$$\text{wp.}Prog_1.(2[s = A] + [s = C]) = 1/2$$

$$\text{wp.}Prog_0.(2[s = A] + [s = C]) = 1$$

The transformer semantics, based on full random variables, is compositional.

A nice proof rule, proved using the transformer semantics:

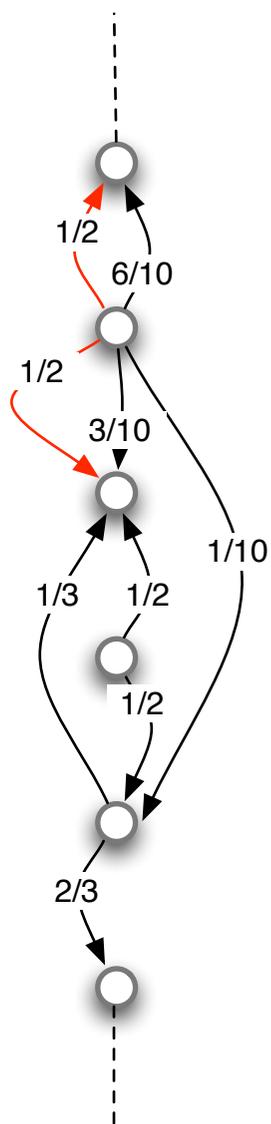
A loop:  $\text{do } G \rightarrow \text{body } \text{od}$

An invariant:  $[G] \times I \leq \text{wp}.\text{body}.I$

Termination condition:  $T \hat{=} \text{wp}.\text{(do } G \rightarrow \text{body } \text{od)}.1$

A rule:  $I \leq T \Rightarrow I \leq \text{wp}.\text{(do } G \rightarrow \text{body } \text{od)}.I$

# The “jumping bean” : specification.



$$[n = N] \leq [\text{wp.jump}.[n \neq N]]$$

The bean must move...

$$[n = N] \leq \text{wp.jump}.[N - K \leq n \leq N + K]$$

(K is a fixed constant.)

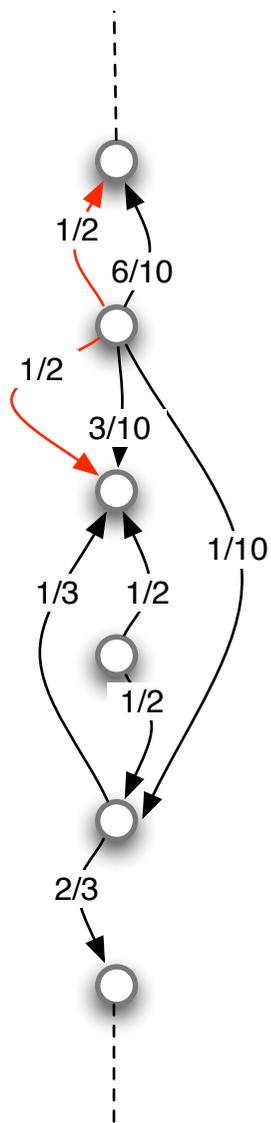
The bean can't move too much...

$$n \leq \text{wp.jump}.n$$

The expected move is at least 0.

# The “jumping bean”.

$$Bean \hat{=} wp.(do (n \leq N) \rightarrow jump \text{ od})$$



$$1 = wp.Bean.[n > N]$$

The bean continues to jump, until it exceeds  $N$ .

The conditions on its behaviour guarantee that it will eventually exceed any bound.

Exercise: use the properties of the transformers to prove this. (Should be about 10 lines of proof.)

## Probability versus nondeterminism:

$$\begin{aligned} & (x := 0 \sqcap x := 1); (y := 0 \text{ }_{1/2} \oplus y := 1) \\ = & (x := 0 \sqcap x := 1); (y := 0) \\ & \text{ }_{1/2} \oplus \\ & (x := 0 \sqcap x := 1); y := 1) \end{aligned}$$

The demon can predict the future.

$$\begin{aligned} & (y := 0 \text{ }_{1/2} \oplus y := 1); (x := 0 \sqcap x := 1) \\ = & y := 0; (x := 0 \sqcap x := 1) \\ & \text{ }_{1/2} \oplus \\ & y := 1; (x := 0 \sqcap x := 1) \end{aligned}$$

The demon can access the past.

## Probability versus nondeterminism:

Smyth powerdomain, for nondeterminism; then the probabilistic powerdomain on top of that.

The demon can predict the future.

Probabilistic powerdomain to make  $EvalS_{\perp}$ , then the Smyth powerdomain to make  $S_{\perp} \rightarrow \mathbb{P}Eval.S_{\perp}$  with a special definition of “;”

The demon can access the past.

Suppose we wanted to prevent the demon from accessing the past, i.e.

$$\begin{aligned} & (y := 0 \oplus_{1/2} y := 1); (x := 0 \sqcap x := 1) \\ = & \begin{array}{c} (y := 0 \oplus_{1/2} y := 1); x := 0 \\ \sqcap \\ (y := 0 \oplus_{1/2} y := 1); x := 1 \end{array} \end{aligned}$$

How would we build a semantic domain justifying this algebraic property?

Suppose we wanted to prevent the demon from accessing the past, i.e.

Use the probabilistic powerdomain to build  $Eval.S_{\perp} \rightarrow Eval.S_{\perp}$ ,  
and then the Smyth powerdomain to build  $Eval.S_{\perp} \rightarrow \mathbb{P}Eval.S_{\perp}$

How would we build a semantic domain justifying this algebraic property?

With this model, we lose some refinements:

$$A \sqcap B \not\sqsubseteq \text{if } G \text{ then } A \text{ else } B$$

In fact, if we did not ban this refinement and we still had the equality we started with, then we would also have to accept this refinement!

$$(x, y := 0, 0)_{1/2} \oplus (x, y := 1, 1) \sqsubseteq x, y := 0, 0$$

# An algebraic “thought experiment”

$(y := 0 \text{ }_{1/2} \oplus y := 1); (x := 0 \sqcap x := 1);$   
if  $(x = y)$  then  $(x, y := 1, 1)$  else  $(x, y := 0, 0)$

$\sqsubseteq$   $(y := 0 \text{ }_{1/2} \oplus y := 1); (\text{if } (y = 0) \text{ then } x := 0) \text{ else } x := 1;$  †  
if  $(x = y)$  then  $Z$  else  $W$

$=$   $(y := 0; x := 0 \text{ }_{1/2} \oplus y := 1; x := 0)$   
if  $(x = y)$  then  $Z$  else  $W$

$=$   $Z$   
 $=$   $x, y := 0, 0$

$(y := 0 \text{ }_{1/2} \oplus y := 1); (x := 0 \sqcap x := 1);$   
 if  $(x = y)$  then  $Z$  else  $W$

$= ((y := 0 \text{ }_{1/2} \oplus y := 1); x := 0) \sqcap ((y := 0 \text{ }_{1/2} \oplus y := 1); x := 1);$  †  
 if  $(x = y)$  then  $Z$  else  $W$

$((y := 0 \text{ }_{1/2} \oplus y := 1); x := 0);$  if  $(x = y)$  then  $Z$  else  $W$   
 $\sqcap$

$((y := 0 \text{ }_{1/2} \oplus y := 1); x := 1);$  if  $(x = y)$  then  $Z$  else  $W$

$(Z \text{ }_{1/2} \oplus W) \sqcap (Z \text{ }_{1/2} \oplus W)$

$= (x, y := 0, 0) \text{ }_{1/2} \oplus (x, y := 1, 1)$

$(x, y := 0, 0) \text{ }_{1/2} \oplus (x, y := 1, 1) \sqsubseteq x, y := 0, 0$

## The “refinement paradox”

Properties of the logic/algebra in the context where “hidden state” is an issue are hard to get right, even when there are no probabilities.

It turns out to be a really hard problem to find a formalisation which behaves properly for refinement

$h$       “High security” variables (are “private”)

$l$       “Low security” variables (are “public”)

“Obviously” we want to make sure that going up the refinement order preserves our security properties.

# The “refinement paradox”

$h \in \{0, \dots, H\}$

Choose some value for  $h$ ,  
but keep it hidden.

$h \in \{0, \dots, H\} \sqsubseteq h := 0$

Refinement doesn't  
preserve the security of  
the choice?

Even refinement of low security variables is a problem.

$$l := 0 \sqcap l := 1 \sqsubseteq \text{if } (h = 0) \text{ then } l := 0 \text{ else } h := 1$$

And distribution through nondeterminism...

$$(P \sqcap Q); R \sqsubseteq P; R \sqcap Q; R$$

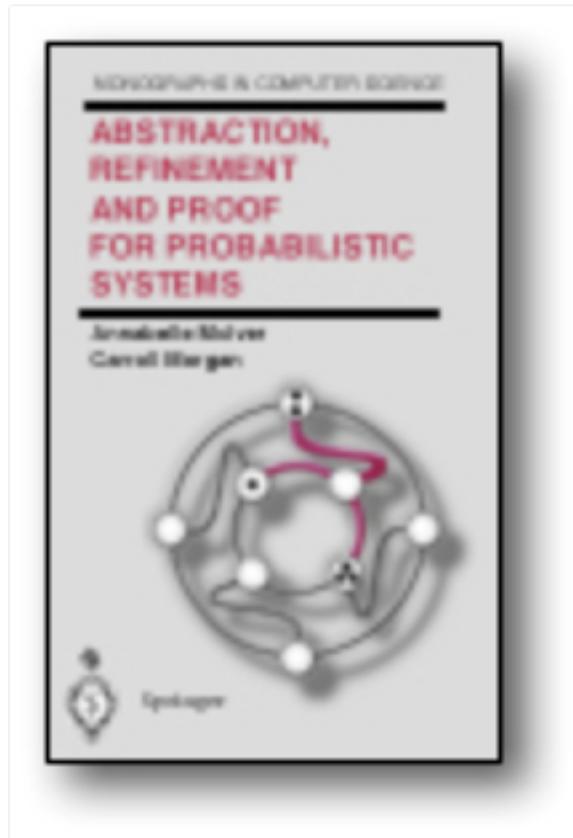
... implies

$$\begin{aligned} & (h := 0 \sqcap h := 1); (l := 0 \sqcap l := 1) \\ = & h := 0; (l := 0 \sqcap l := 1) \sqcap h := 1; (l := 0 \sqcap l := 1) \\ \sqsubseteq & h := 0; l := 0 \sqcap h := 1; l := 1 \end{aligned}$$

## Resolving the paradox

Building a model with “hidden probabilities”, based on the ideas of the final example domain provides the right insights to build a model and logic for “refinement” and knowledge:

- “The Shadow Knows: Refinement of Ignorance in Sequential Programs”, CC Morgan, to appear, Science of Computer Programming



[http://www.cse.unsw.edu.au/~carrollm/probs/  
bibliography.html](http://www.cse.unsw.edu.au/~carrollm/probs/bibliography.html)