Doha, Qatar
The International School
on Research Impact Assessment
"Learning to assess research with the aim to optimise returns"

INFERENCE, PROGRAMME THEORY, FRAMEWORKS AND TRADE OFFS

Building it up

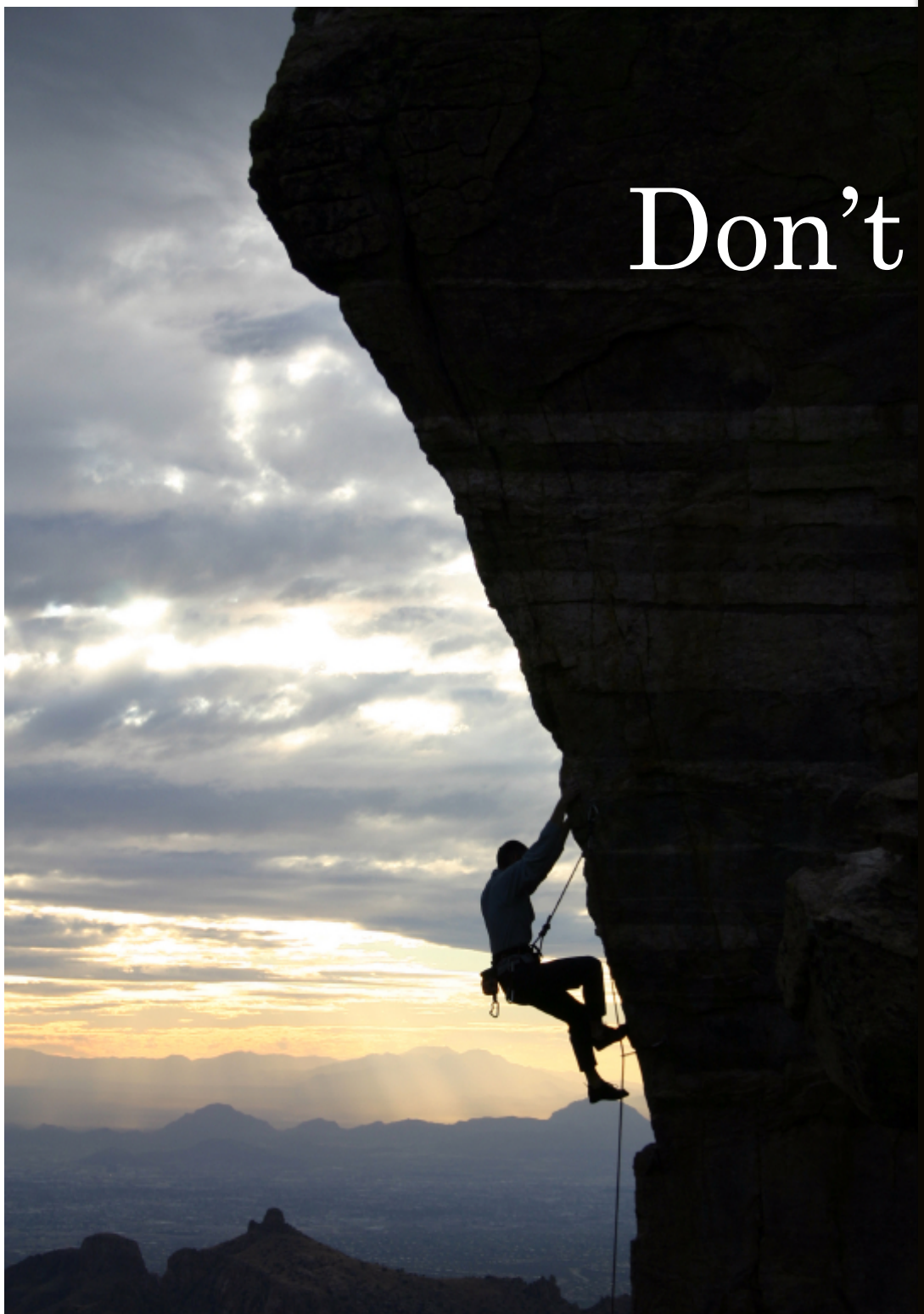
STEVEN WOODING
RAND EUROPE

NOVEMBER 10, 2015

Qatar National Research Fund
Member of Qatar Foundation
Alberta Innovates Health Solutions
Agència de Qualitat i Avaluació Sanitàries de Catalunya
KING'S College LONDON
ASPIRE Academy for Sports Excellence
Sidra Medical and Research Center
Member of Qatar Foundation
researchfish®
RAND EUROPE

Learning outcomes

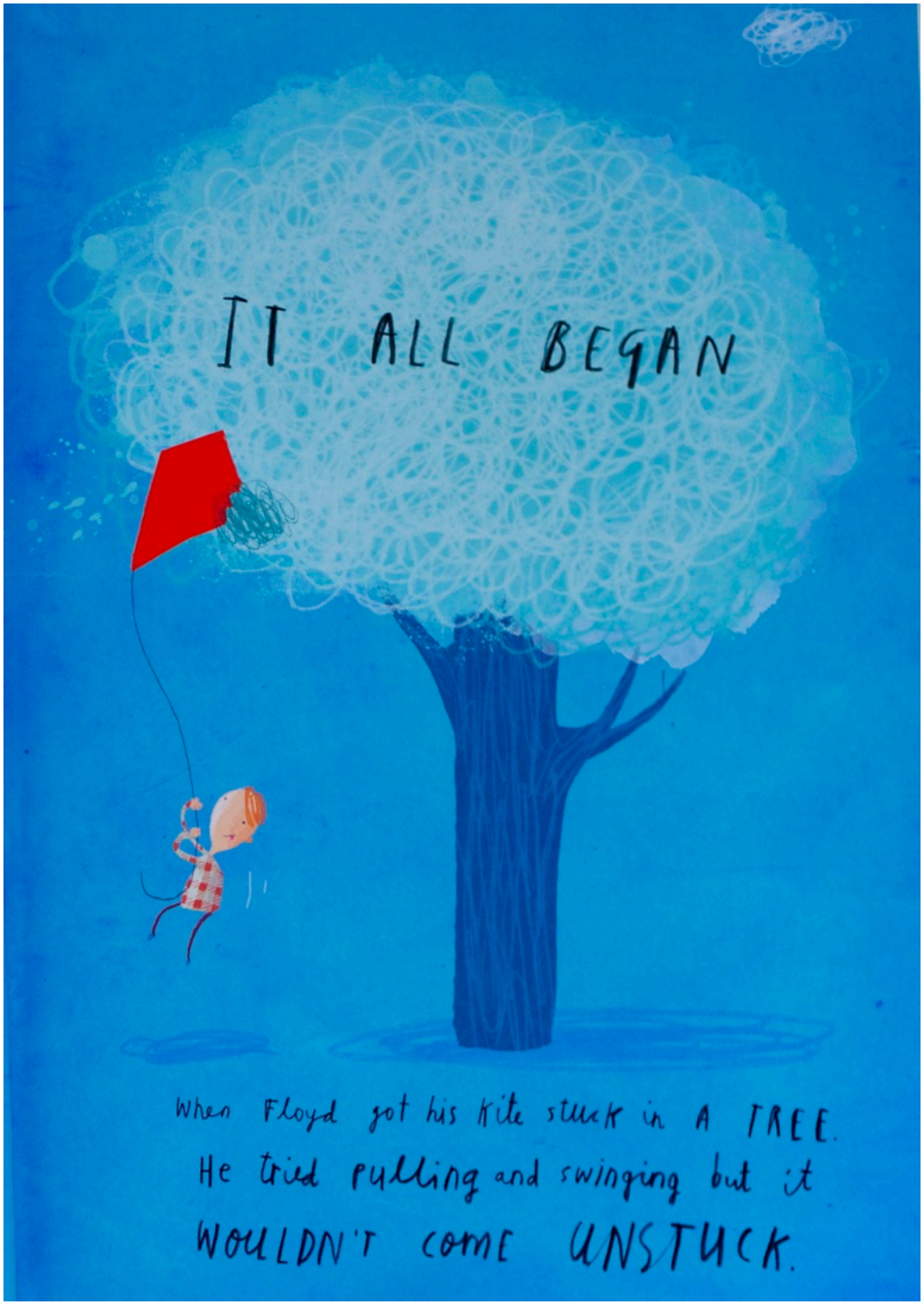
- Understand different approaches to causal inference
- Appreciate experimental and theory based approaches to evaluation and how to combine them
- Understand how to use logic modelling to develop a theory of change for a programme
- Be exposed to the variety of frameworks
- Appreciate the trade offs implicit in research evaluation



Don’t Panic

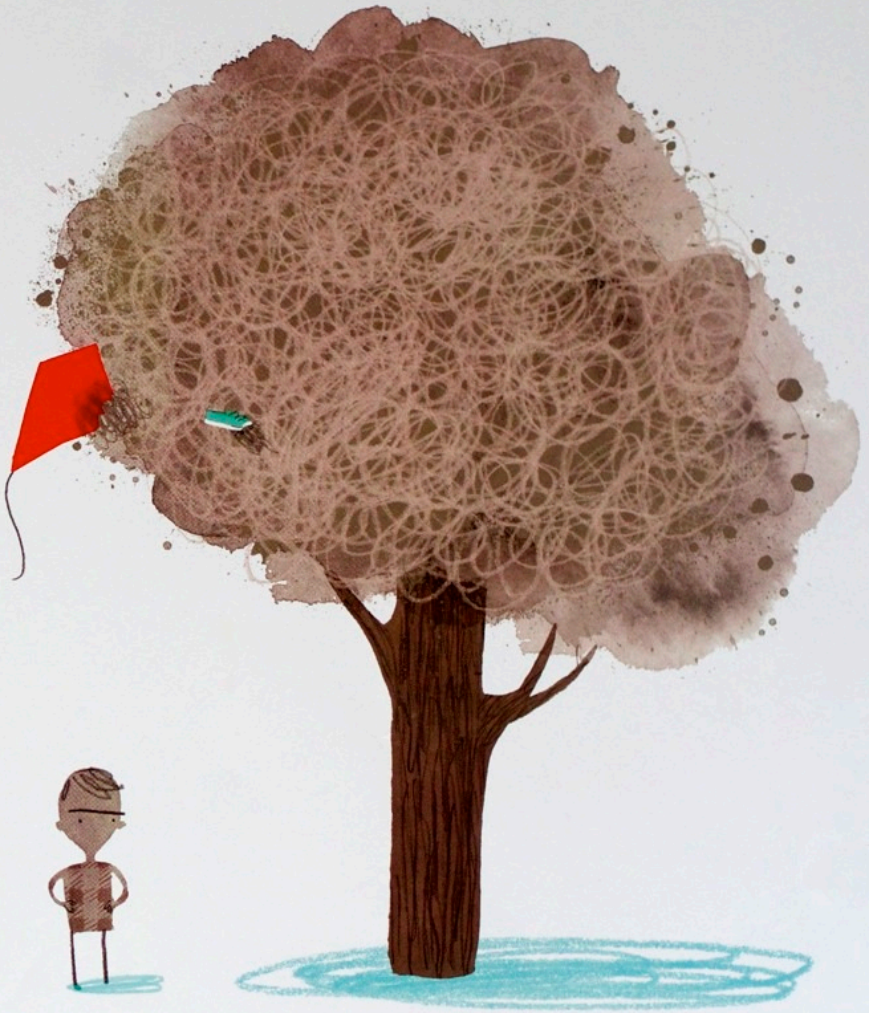
Causal inference


How do we know about the world?


IT ALL BEGAN
When Floyd got his kite stuck in A TREE.
He tried pulling and swinging but it
WOULDN'T come UNSTUCK.

The trouble
REALLY began

when he threw his
FAVOURITE SHOE
to knock the kite loose...



... and THAT got stuck too!

Two types of causal inference

- Theory based evaluation
  - Understanding the process by which x causes y
- Examples
  - case studies
  - expert interviews

- Experimental/ comparative
  - When x happens y always happens even when other things change
- Examples
  - randomised trials
  - natural experiments
  - case control
  - pre/post
  - econometrics

COMPARISON

	Theory based	Experimental
Data requirements	Few cases understood in detail	Many cases matched for important characteristics and diverse in others
Internal validity (resistance to bias)	Low - subject to preconceptions of investigator/experts	High - can overcome bias by testing against data
External validity (ability to generalise to other contexts)	Strong external validity - can compare other situations against evaluated context	Weak - unclear which are most important factors of success

Paper airplane example 1

- Which paper airplanes which fly further?
  - Big or small
- Theory based
  - understand how paper airplanes fly
  - smaller airplanes are lighter
- What about cardboard planes?
  - Knowledge of context: driving factor is weight

Paper airplane example 2

- Which paper airplanes which fly further?
  - Big or small
- Experimental
  - throw them

- What about cardboard planes?
  - Do experimental results apply to cardboard planes?

MOST RESEARCH EVALUATION IS A MIX

- Final outcomes hard to track and attribute – so use theory to work out intermediate outcomes
- Then use experimental\comparative design to see what approach produces the most intermediate outcome

SHOWCASE EVALUATION

- The Wellcome Trust's Showcase scheme aimed to fund high risk, high return research
- Experimental-Case Control
- Theory: high risk research is a good thing
- Method
  - Project descriptions re-written to make them 'scheme agnostic' and reviewed applicants to ensure accuracy
  - Project descriptions compared against control sample of normal project grants of similar size
  - Showcase grants perceived to be more 'risky', 'novel', 'speculative', 'adventurous' and 'innovative' by expert panel members

He threw up his other shoe to knock down his FAVOURITE shoe... and, UNBELievably, that got STUCK as well.

In order to knock down his other shoe,

Floyd fetched Mitch.

CATS get STUCK in trees all the time, but this WAS GETTING RIDICULOUS.

Floyd fetched a ladder.

He was going to sort this out once and FOR ALL...

Recap

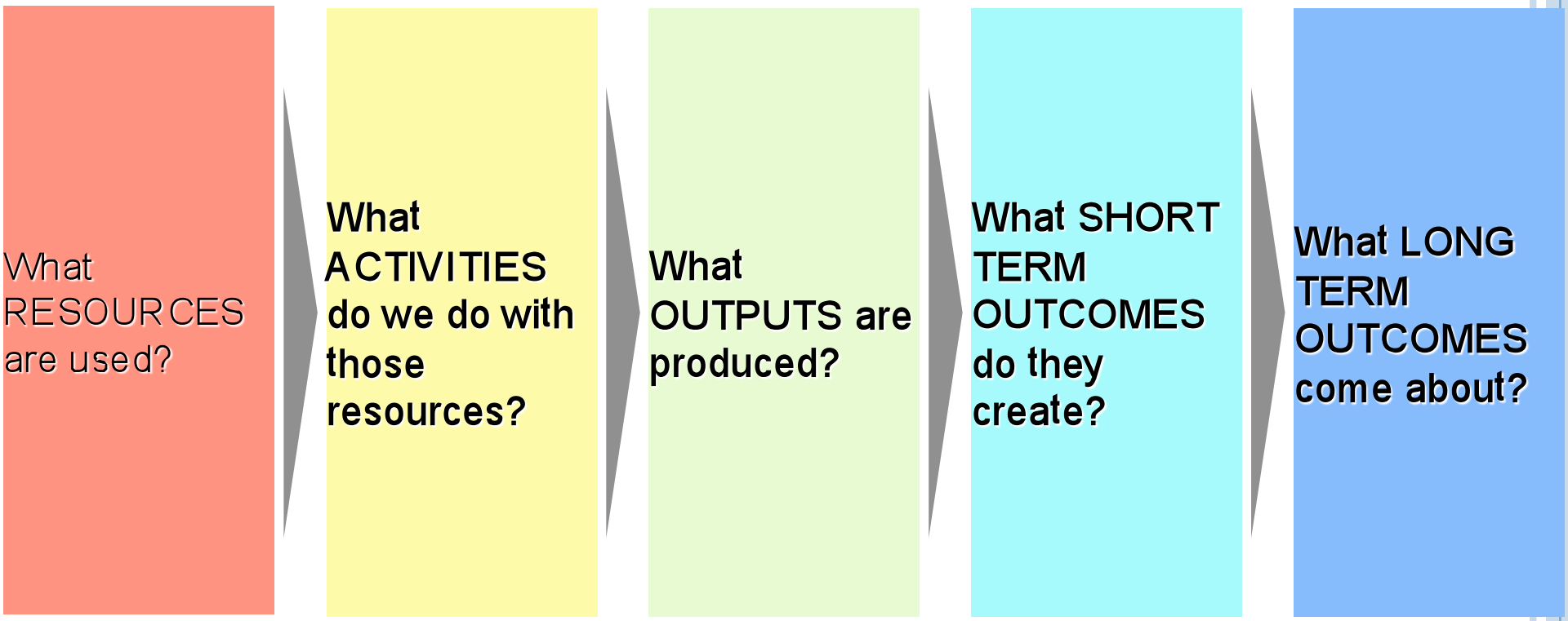
- How do we know stuff?
  - Theory based
  - Experimental/comparative
  - We need both

Programme theory

How do you think your programme works?

Programme theory is a story
Logic models are a picture book

LOGIC MODELS



- A logic model is a flow chart of the programme
- It describes the assumed logical (causal) relationships among programme elements and the problem to be solved

CONSTRUCTING THE LOGIC

What RESOURCES are used?

What ACTIVITIES do we do with those resources?

What OUTPUTS are produced?

What SHORT TERM OUTCOMES do they create?

What LONG TERM OUTCOMES come about?

IF X THEN Y

HOW IS X PRODUCED FROM Y

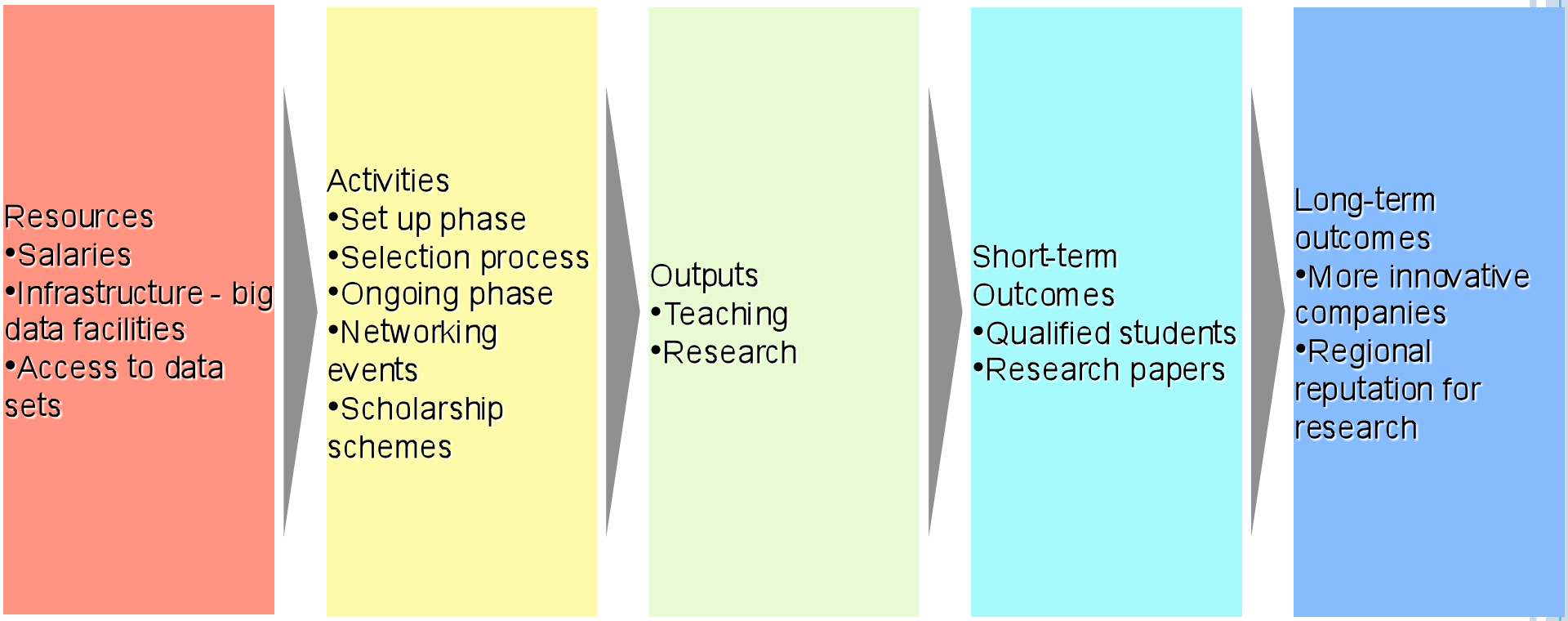
Who finds logic models useful?

- Evaluators
  - Helps understand the programme
  - Guides measurement and data collection
  - Focuses on outcomes and impacts
- Stakeholders
  - Promotes communication of results
  - Provides insight into the key elements of a program
  - Supports face validity
- Across both groups
  - Consensus building and common language

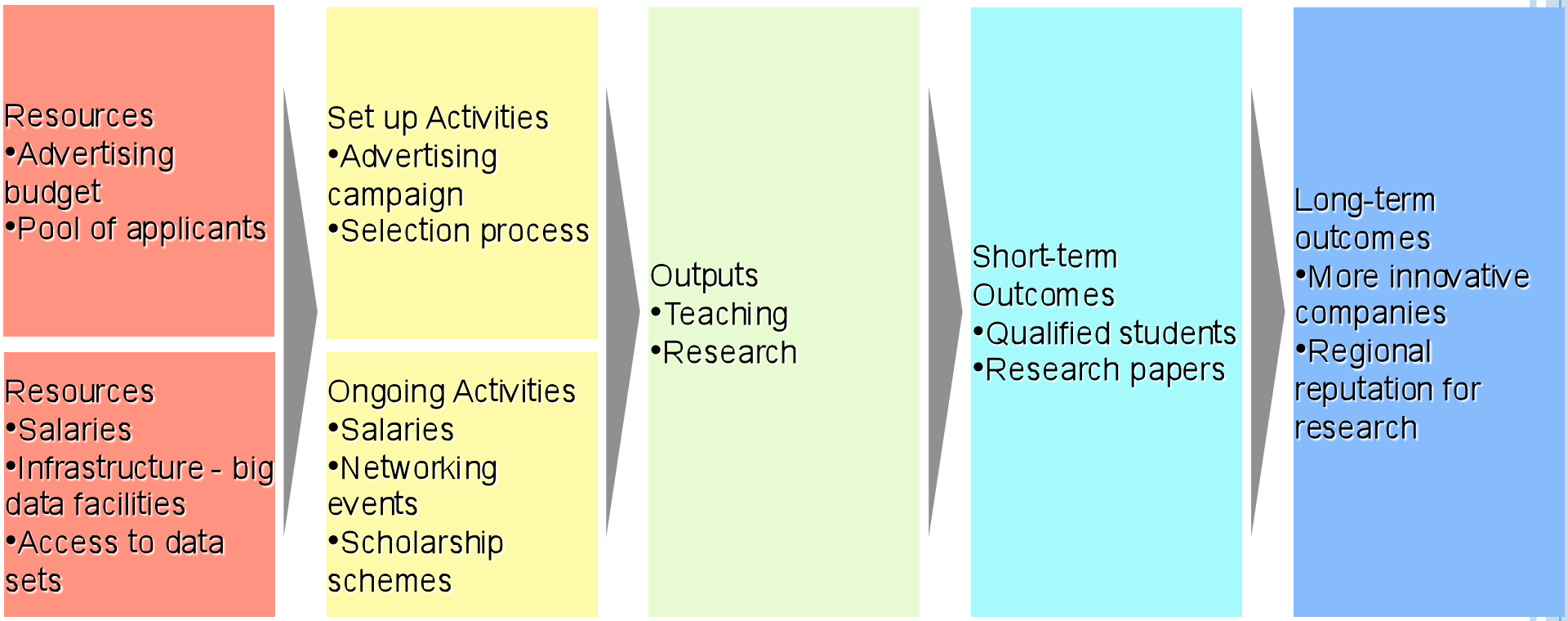
DATA SCIENCE FELLOWSHIPS – AN EXAMPLE

- A new fellowship programme is designed to increase capacity in data science in a region
  - Fellowships will attract excellent researchers to the region
  - They will
    - do research that builds the reputation of regional institutions
    - teach students, who will go on to become data scientists
  - Some of those students will move out of research and into industry leading to economic growth

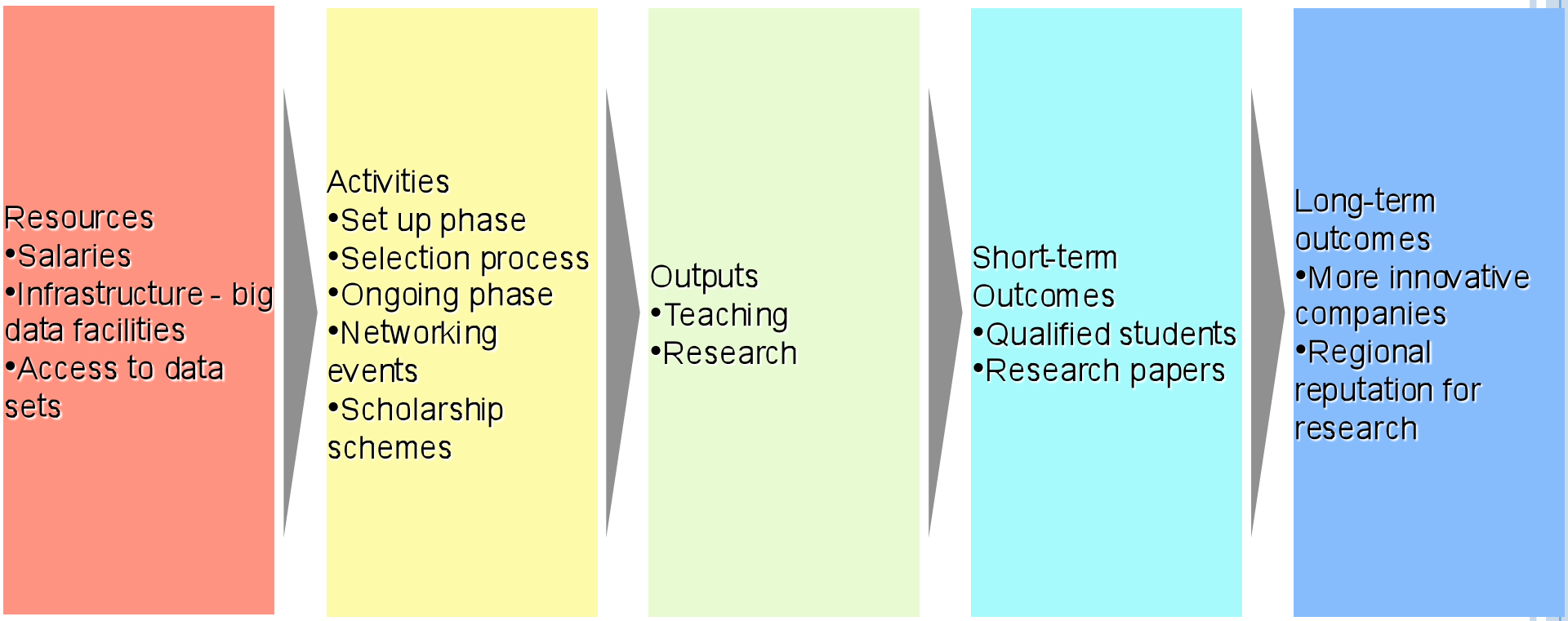
EXAMPLE LOGIC MODEL



REFINING DEPENDENCIES



BEWARE OF OUTSIDE INFLUENCES



Qualified students leave region because they cannot afford house prices

Dangers of logic models

- Limitations
  - Represents intention, not reality
  - Doesn't address: Are we doing the right thing?
  - Focuses on expected outcomes
  - Challenge of causal attribution
  - Too linear
- Cautions
  - Can become too complex and time consuming
  - Don't fall in love with your logic model
  - Becomes fixed rather than flexible and dynamic

CATS get STUCK in trees all the time, but this WAS GETTING RIDICULOUS.

Floyd fetched a ladder.

He was going to sort this out once and FOR ALL...

... and up he threw it.

I'm sure you can guess what happened.

The ladder was borrowed
From a neighbour and
would DEFINITELY
need to be put back before
anyone noticed...

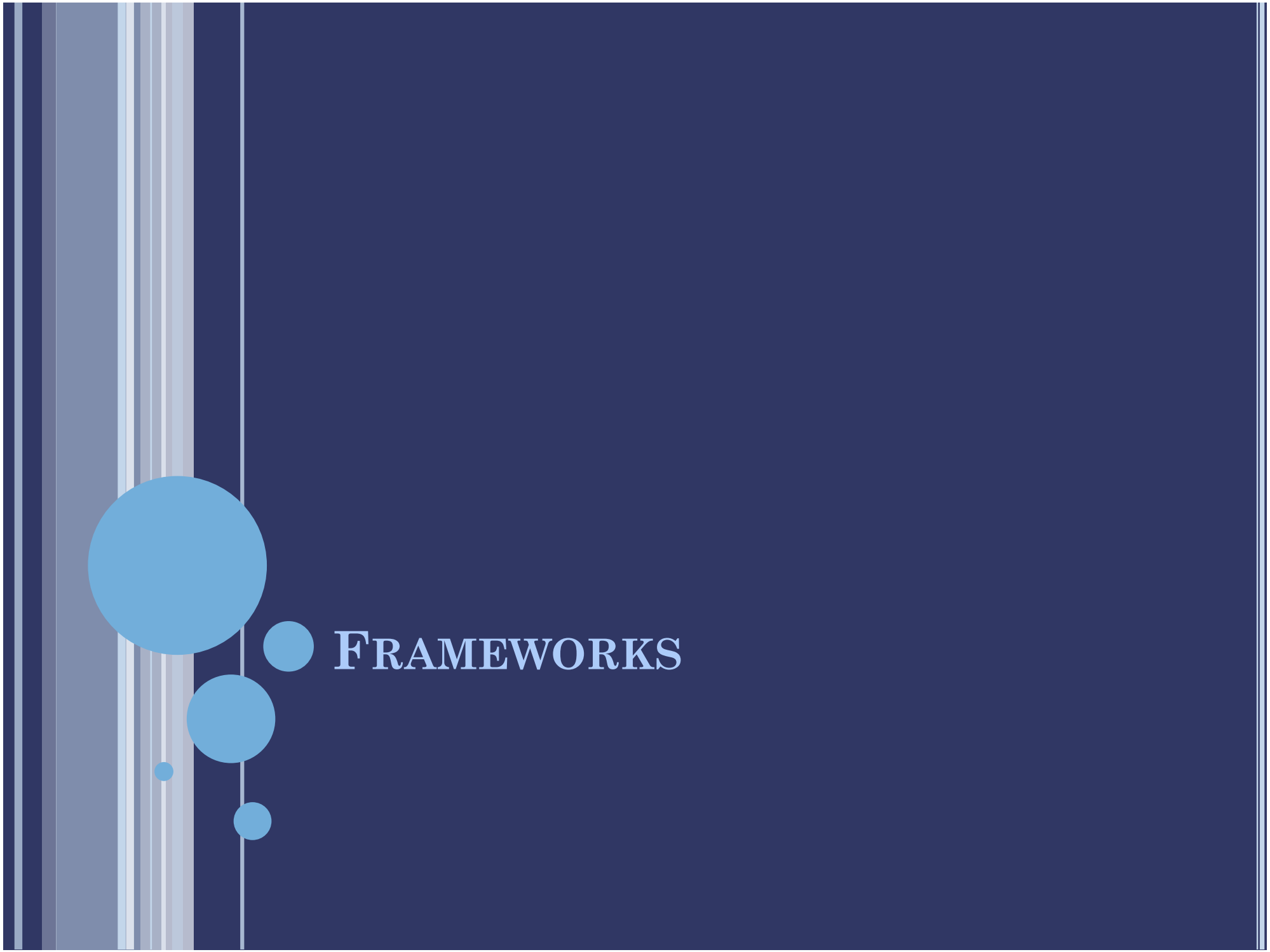
and in order to do so,
Floyd FLUNG a BUCKET of
PAINT at it.

And wouldn't you know...
the BUCKET of paint got STUCK.

Then Floyd tried...

Recap

- Looked at how we know stuff
  - Theory based
  - Experimental/comparative
  - We need both
- Used logic models to explore how our programme works


Frameworks

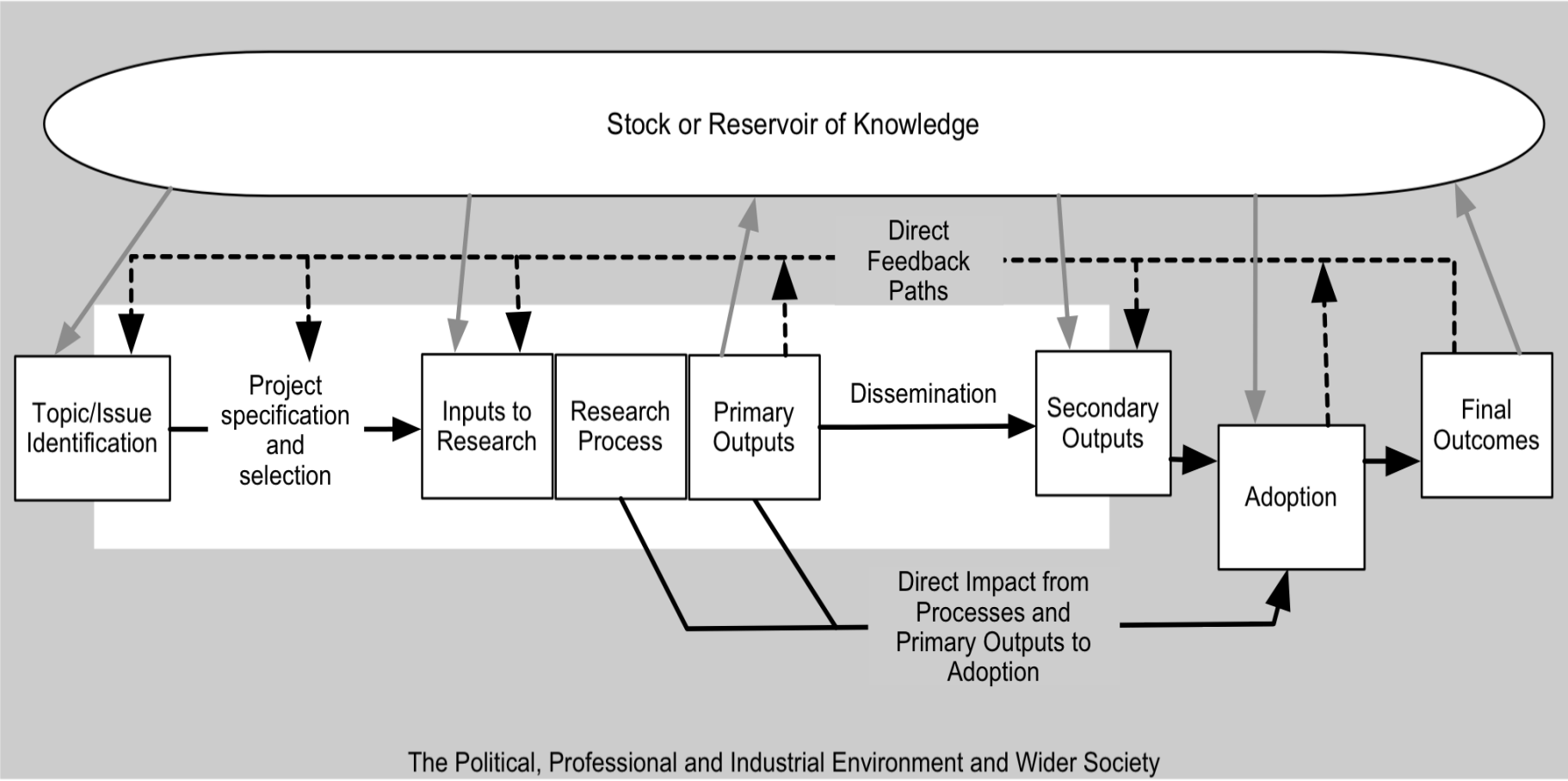
THE RETURNS FROM ARTHRITIS RESEARCH

- 16 case studies based on individual research grants, selected to mirror variety of ARC funding from 15 years previously
  - Project n=6, Programme n=3, Fellowship n=3, Institute n=4
- Wanted to understand the different outputs and outcomes and what led to those differences
- Data sources
  - Archival document review
  - Interviews with Principal Investigators and other researchers
  - Review of published outputs
  - Bibliometric tracing

The Payback Framework

+

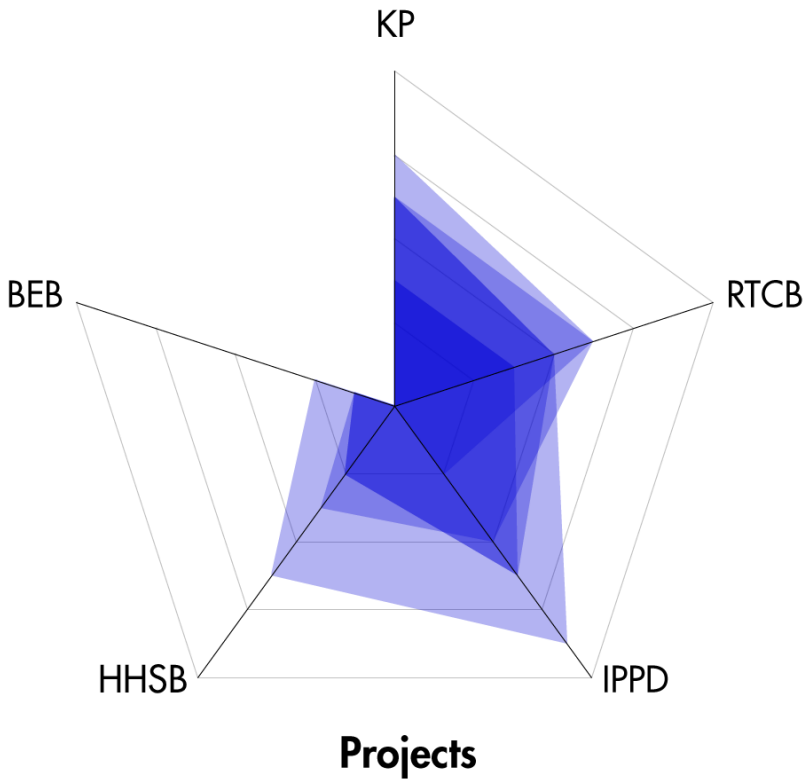
Payback Model



Payback Categories

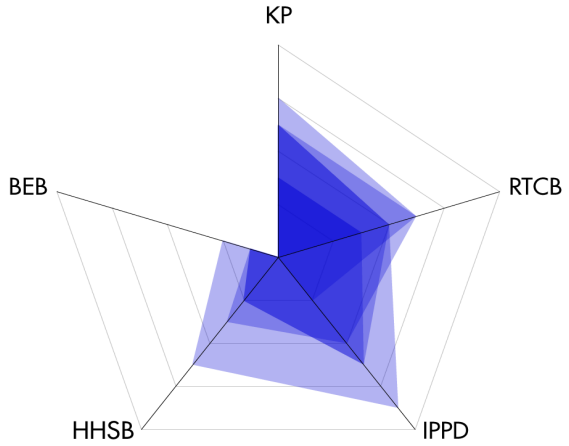
- Knowledge production
- Research Targeting, Capacity Building
- Informing Policy or Product Development
- Health and Health Sector Benefits
- Broader Economic Benefits

Overlaying Profiles

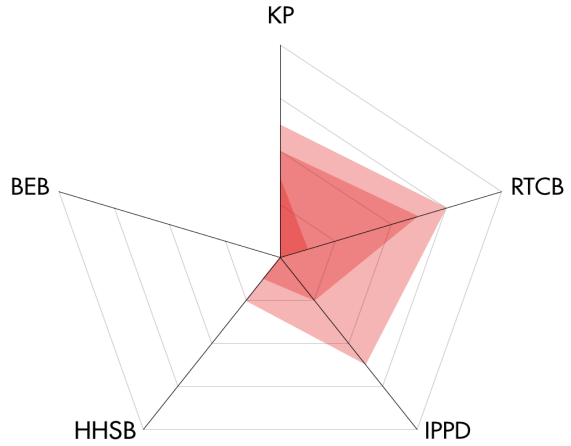


- **Knowledge Production**
- **Research Targeting and Capacity Building**
- **Informing Policy and Product Development**
- **Health and Health Sector Benefits**
- **Broader Economic Benefits**

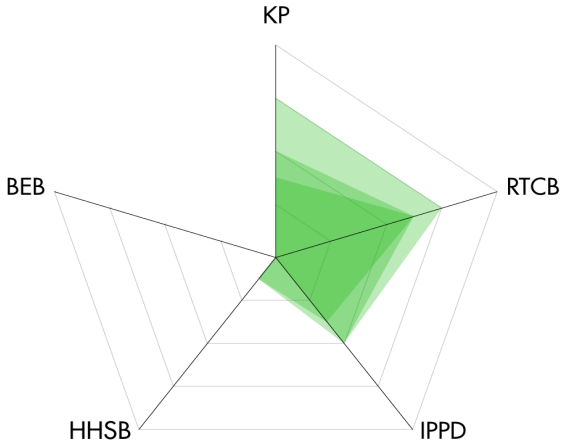
Analysis: Project grants do well



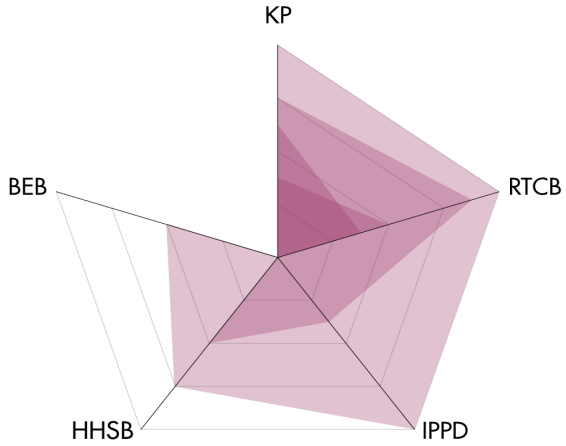
Projects
n = 6



Programmes
n = 3



Fellowships
n = 3



Institutes
n = 4

Returns to Arthritis Research

- Improving policy
  - Provided support for maintaining project grant funding
  - Intended and unintended flexibility is used advantageously —ie, reassurance about increased flexibility
- Accountability and advocacy
  - Demonstrated diversity and extent of payback
  - Case studies

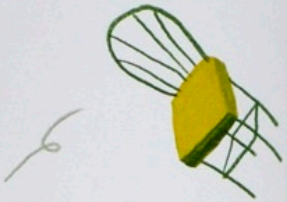
The ladder was borrowed
From a neighbour and
would DEFINITELY
need to be put back before
anyone noticed...

and in order to do so,
Floyd FLUNG a BUCKET of
PAINT at it.

And wouldn't you know...
the BUCKET of paint got STUCK.

Then Floyd tried...

a duck to
knock down the
bucket of paint...



a chair
to knock down
the duck...

his friend's bicycle
to knock down
the chair...

the kitchen sink
to knock down
his friend's bicycle...

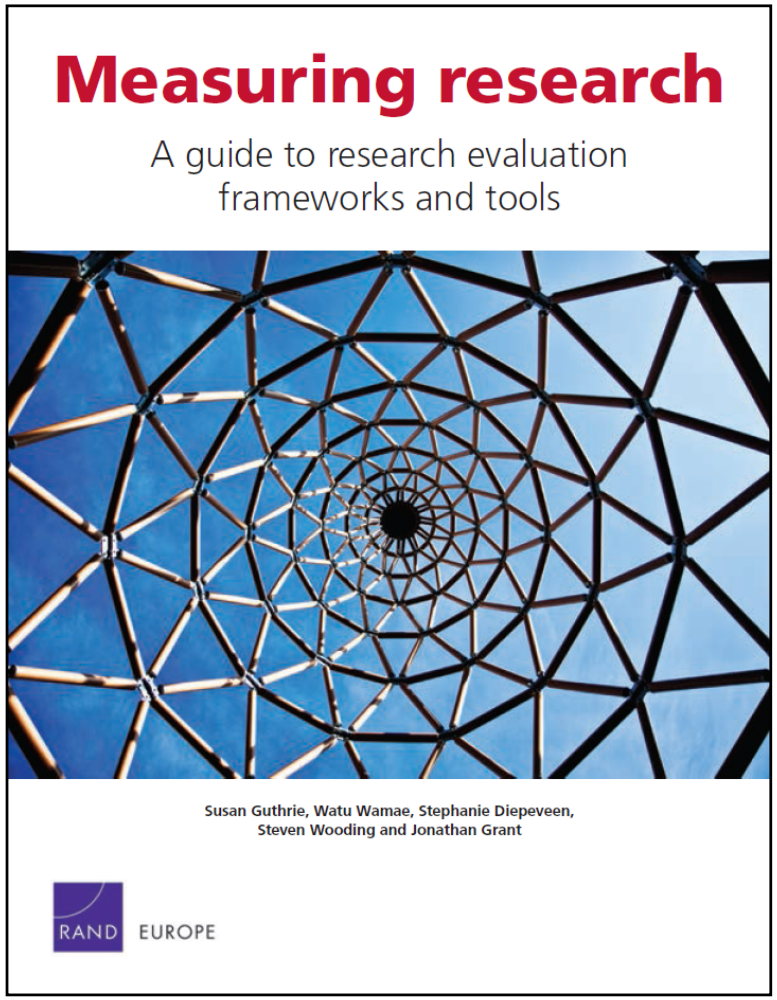
their front door to knock
down the kitchen sink...

Recap

- Looked at how we know stuff
  - Theory based
  - Experimental/comparative
  - We need both
- Used logic models to explore how our programme works
- Introduced the Payback framework as exemplar framework

MEASURING RESEARCH

- A survey of research evaluation frameworks that are in use
- Summary information on key methods
- Discussion of the challenges and trade-offs in evaluating research
- A 'how-to guide' to evaluating research
- Promote discussion around the science of science



THE VARIETY OF FRAMEWORKS…


Canadian Academy of Health Science (CAHS), CA


Excellence in Research for Australia (ERA), AU


Research Excellence Framework (REF), UK


STAR METRICS, US


National Institute of Health Research (NIHR) Dashboard, UK


Productive Interactions, EU

ERA AND THE REF

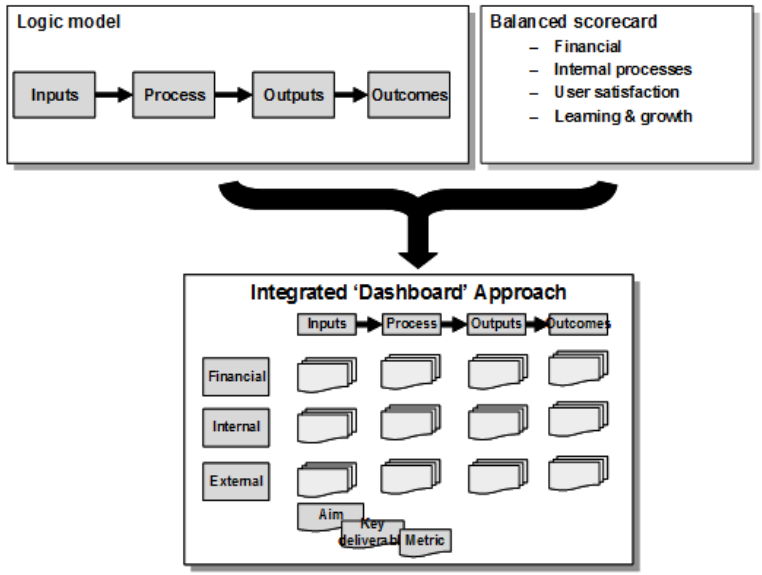
	Emphasis	Uniformity
ERA	Quantitative indicators	Two different approaches for different disciplines
REF2014	Expert review by panels	Same approach for all disciplines

STAR METRICS – STAGE ONE

- **How many jobs** created by US assessment of federally funded science research
- Metrics based approach
  - Mines existing institutional administration data sets
  - Minimise burden on researchers
  - Participation voluntary
- Stage Two
  - Wider understanding of research conducted, including indicators of impacts on economic growth, workforce outcomes, scientific knowledge, social outcomes.

NIHR Dashboard – Overview

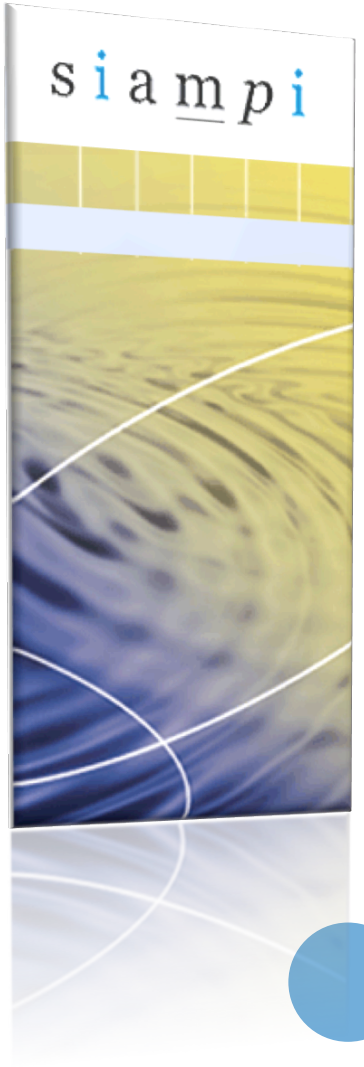
- **Management tool**:
  - Supports strategic decision-making by providing performance measures on a regular basis



- Small but balanced set of tailor-made indicators
- Balanced view of performance andlow burden of data collection
- Combines logic model with balanced scorecard, collecting programme-level data
  - on inputs, process, outputs and outcomes (logic model)
  - for financial, internal process and user satisfaction (balanced scorecard)

Productive Interactions - Overview

- Uses 'Productive Interactions' (PI) as a proxy for impact
- Premise: knowledge develops and impact is achieved through a series of interactions between researchers and society
- Modelled as a two way process, three types of Productive Interactions:
  - Direct personal contacts
  - Indirect interaction
  - Financial interaction
- **Outcomes used for reflection and learning**, not external or comparative assessment



the FAMILY car
to knock down
their front door...

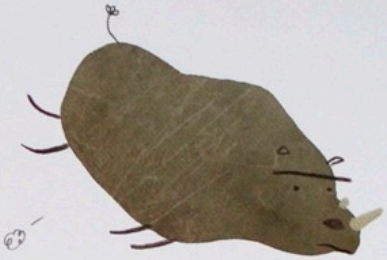
the
MILKMAN
to knock down
the Family car...

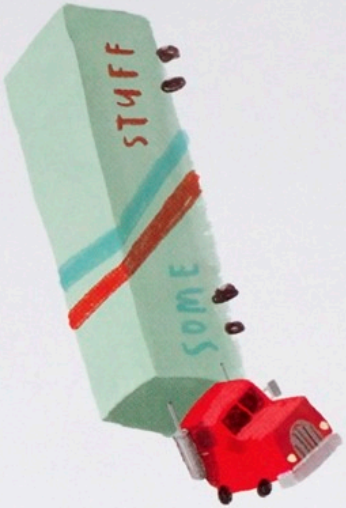
an Orang-utan to KNOCK DOWN
the milkman, who surely had
somewhere else to be...

a small boat
to knock down
the orang-utan...

a BIG
BOAT
to knock
down the
small
boat...

a ~~B~~ RHINOCEROS
to knock down the
BIG boat...





a long-distance lorry
to knock down the
rhinoceros...

the HOUSE across
the street to
knock down the
long-distance lorry...


A LIGHTHOUSE to KNOCK DOWN the house no longer across the street...
HI! What are you doing?
a curious whale, in THE WRONG PLACE at THE WRONG TIME, to knock down the lighthouse...

and they

ALL

GOT

STUCK.

Recap

- Looked at how we know stuff
  - Theory based
  - Experimental/comparative
  - We need both
- Used logic models to explore how our programme works
- Introduced the Payback framework as exemplar framework
- Reviewed the diversity of frameworks

Tensions in Research Impact Assesment

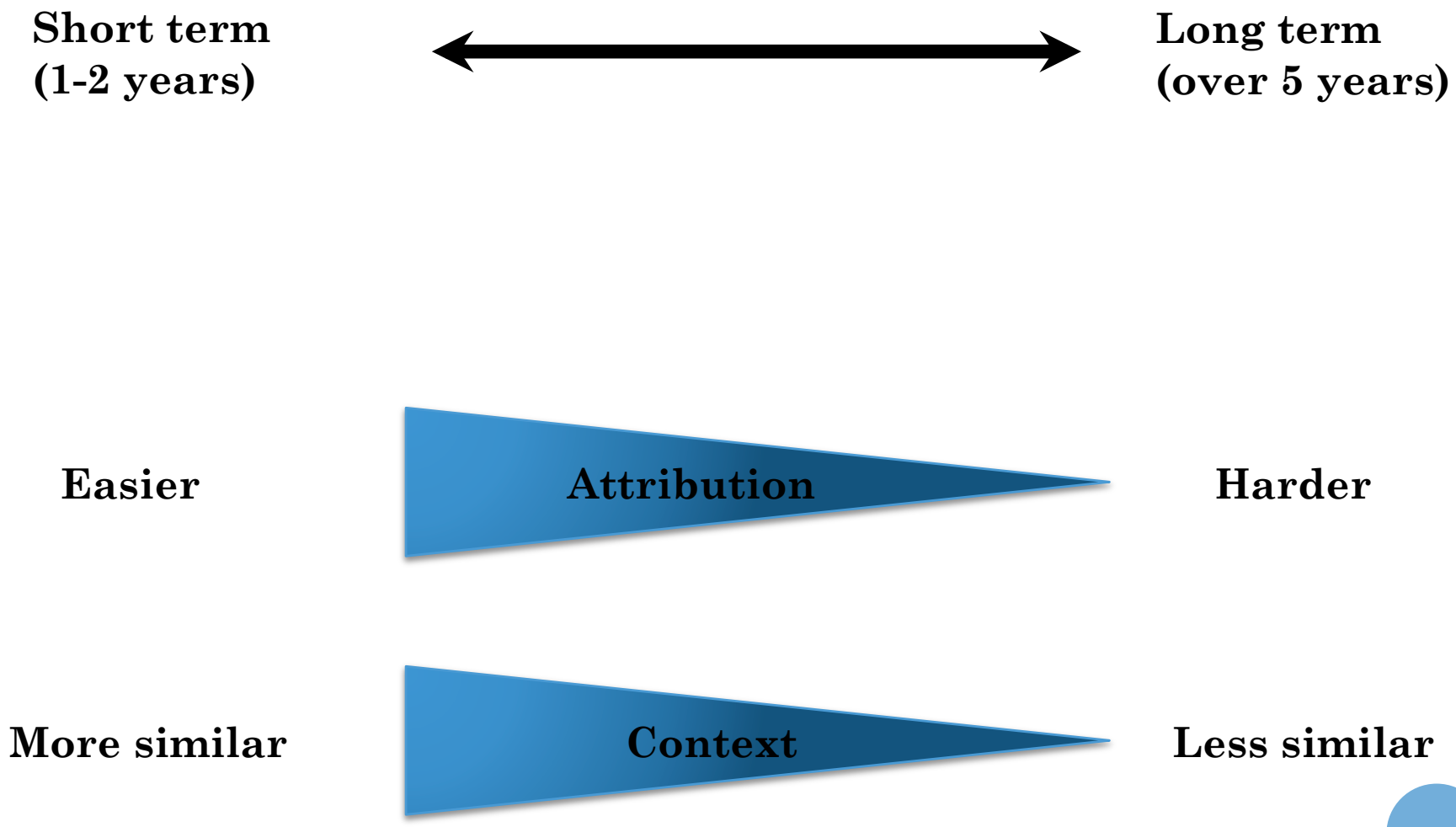
Key Tensions

Short term (1-2 years)	⟷	**Long term (over 5 years)**
Depth	⟷	**Breadth**
Flexibility	⟷	**Comparability**
Improvement	⟷	**Assessment**

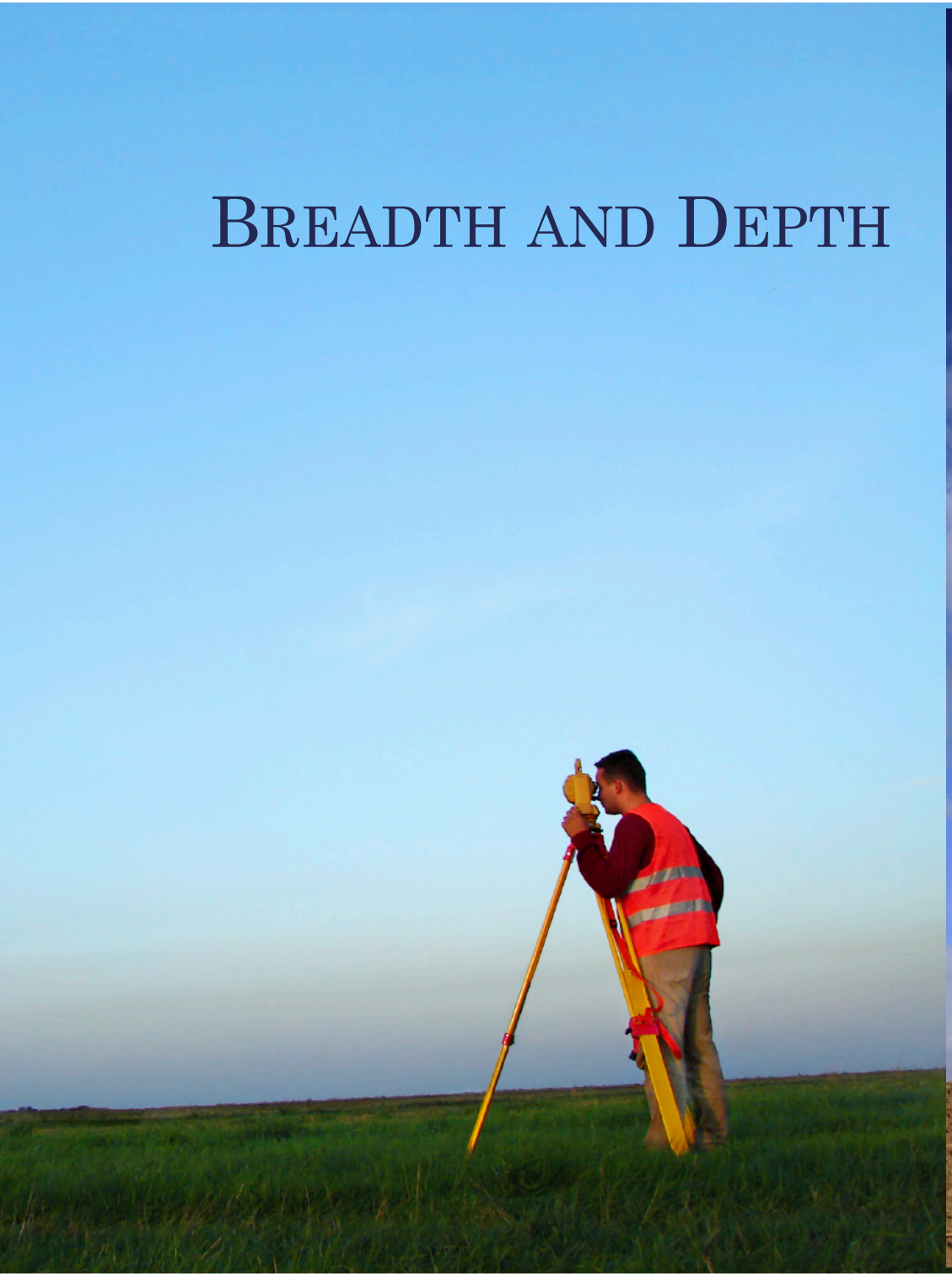
TIMESCALE FOR EVALUATION

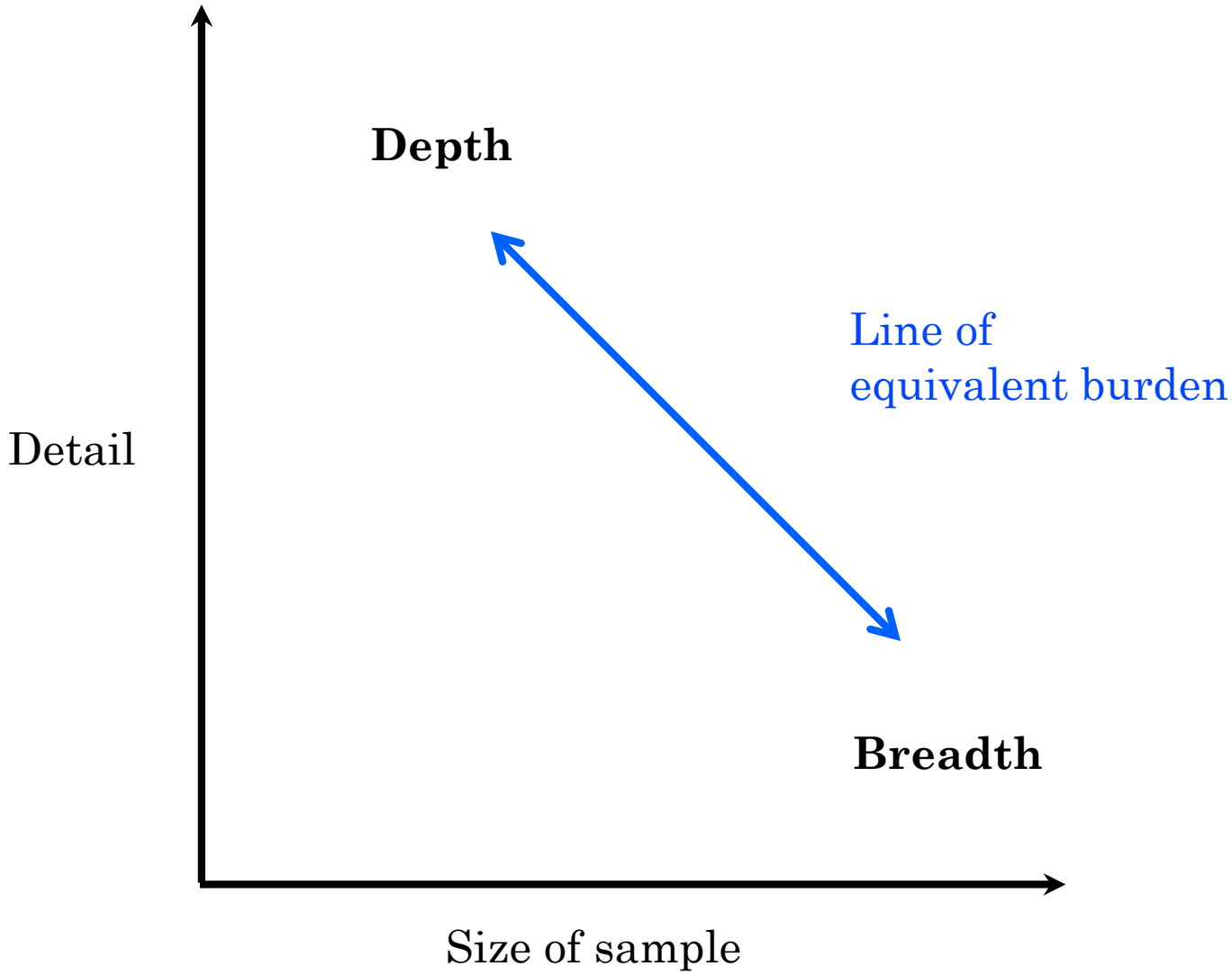
Short term (1-2 years) ⟷ **Long term (over 5 years)**

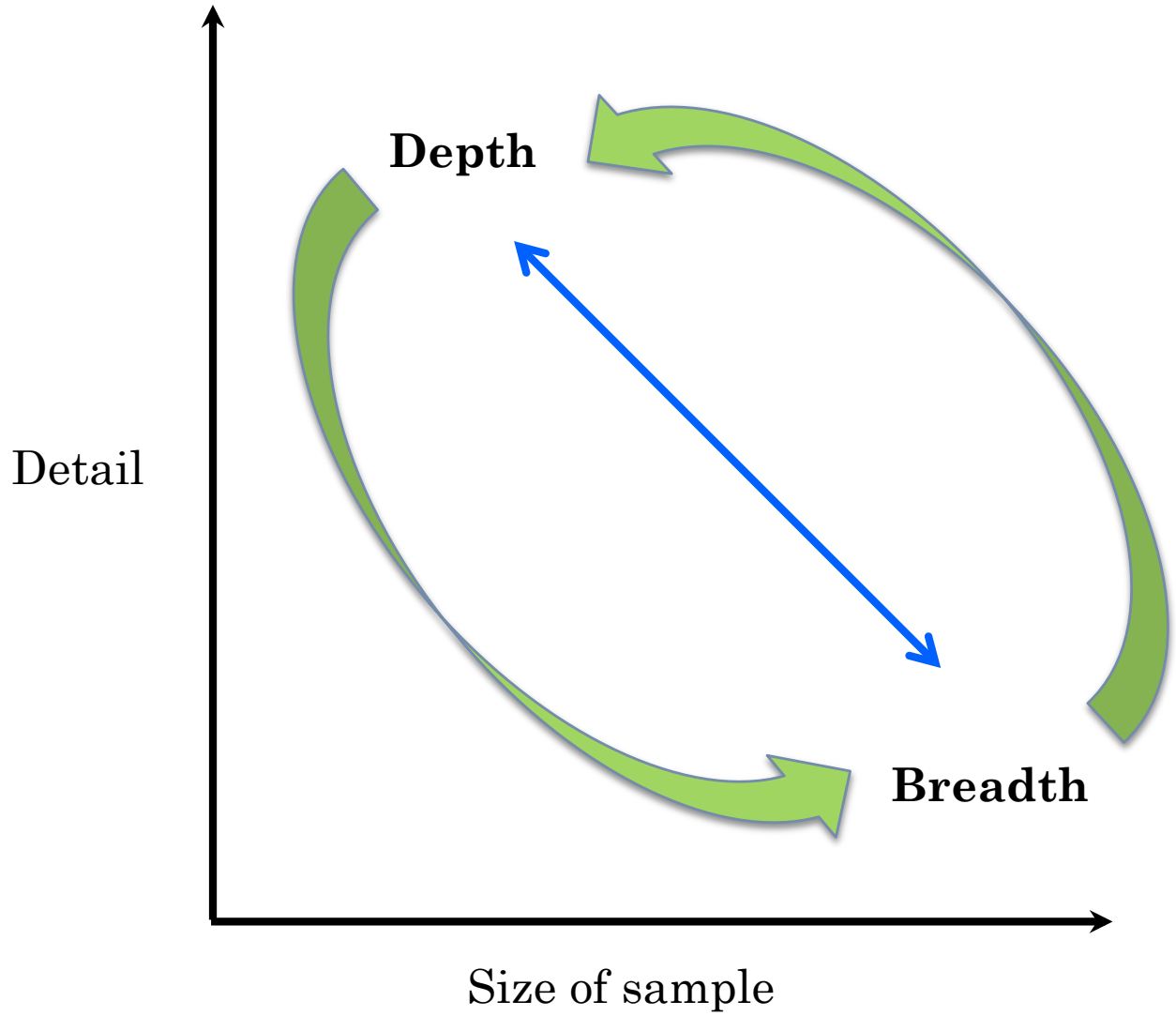
Process and outputs ⟷ **Outcomes**

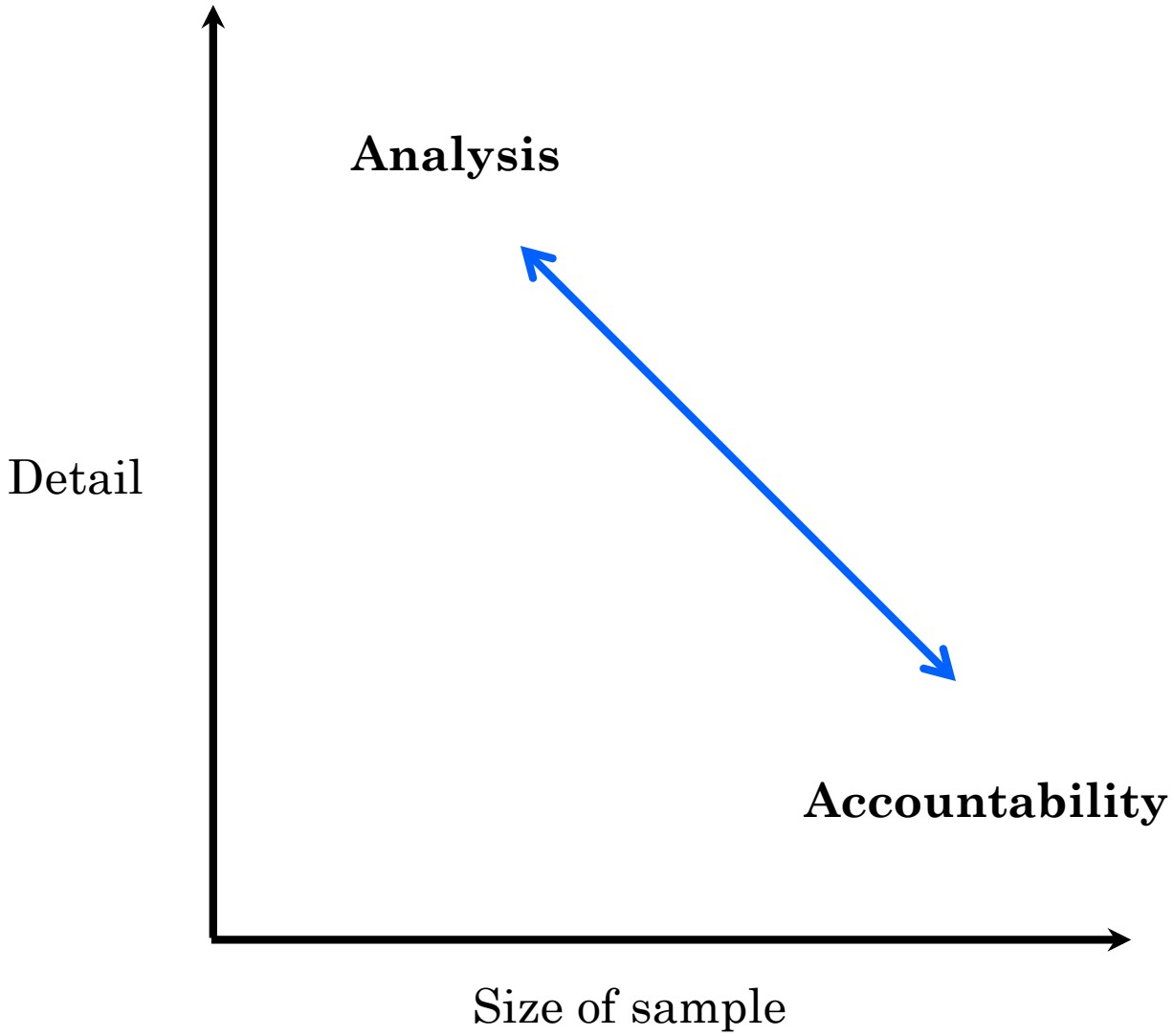

Short term
(1-2 years)
Long term
(over 5 years)
Easier
Attribution
Harder
More similar
Context
Less similar

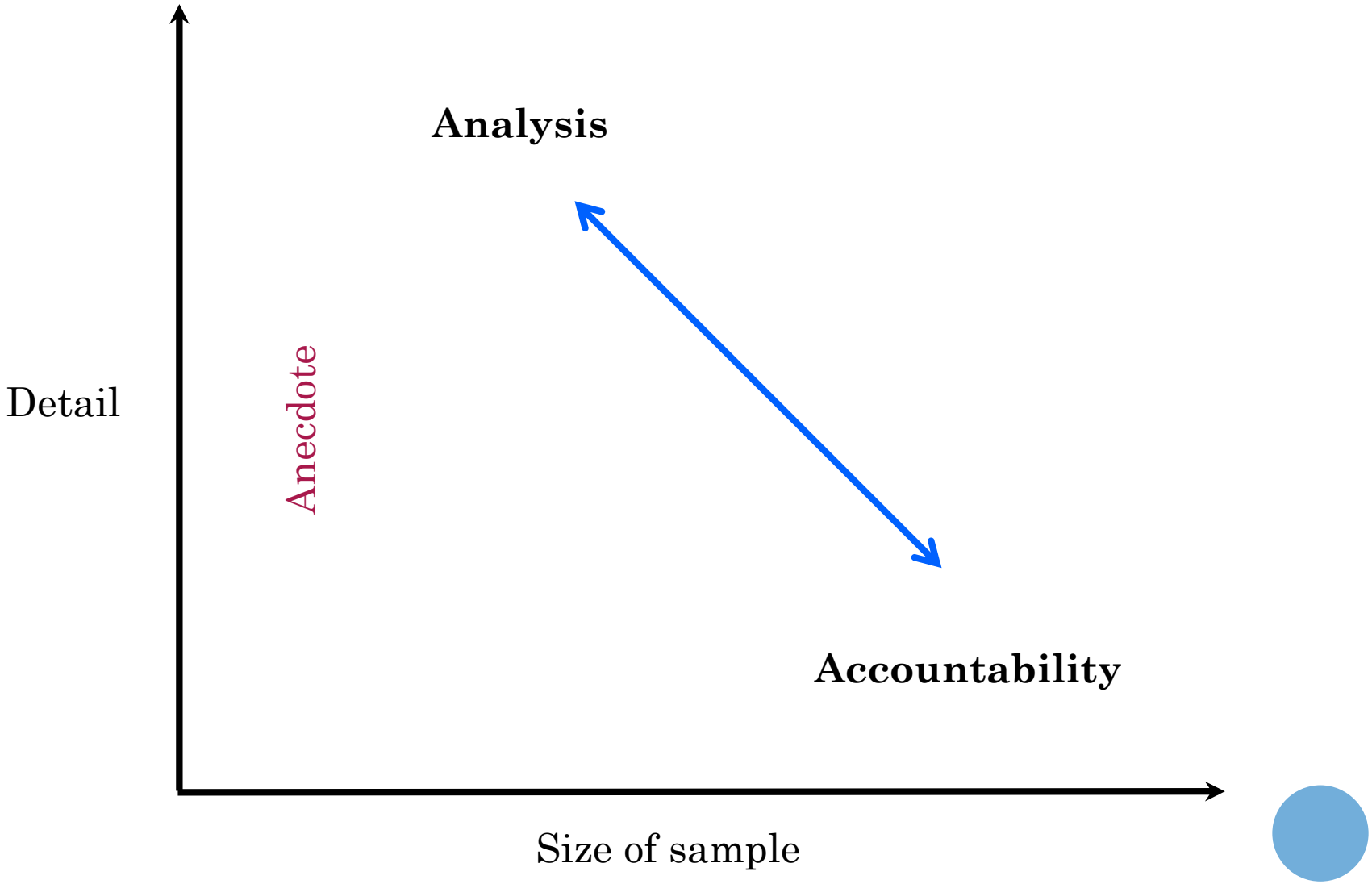
Breadth and Depth




Depth
Line of
equivalent burden
Detail
Breadth
Size of sample


Depth
Detail
Breadth
Size of sample


Analysis
Detail
Accountability
Size of sample


Analysis
Detail
Anecdote
Accountability
Size of sample

Burden

Researchers vs Secretartia

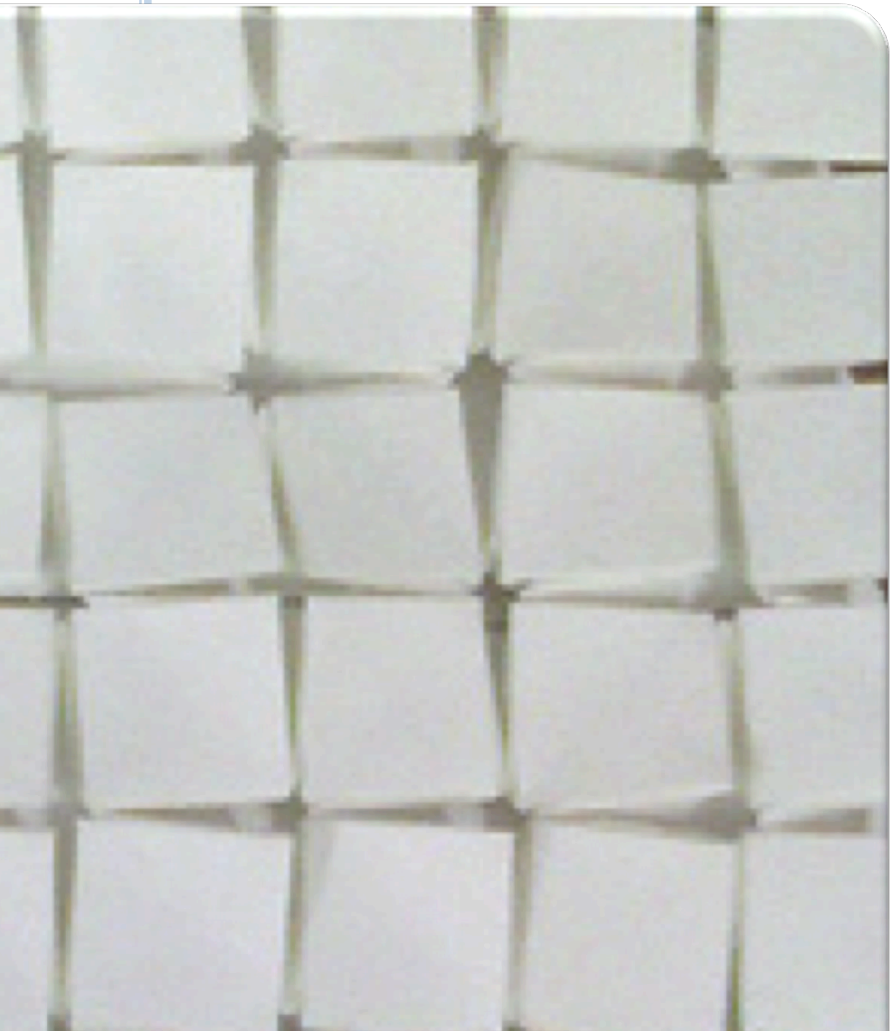
Initial vs ongoing

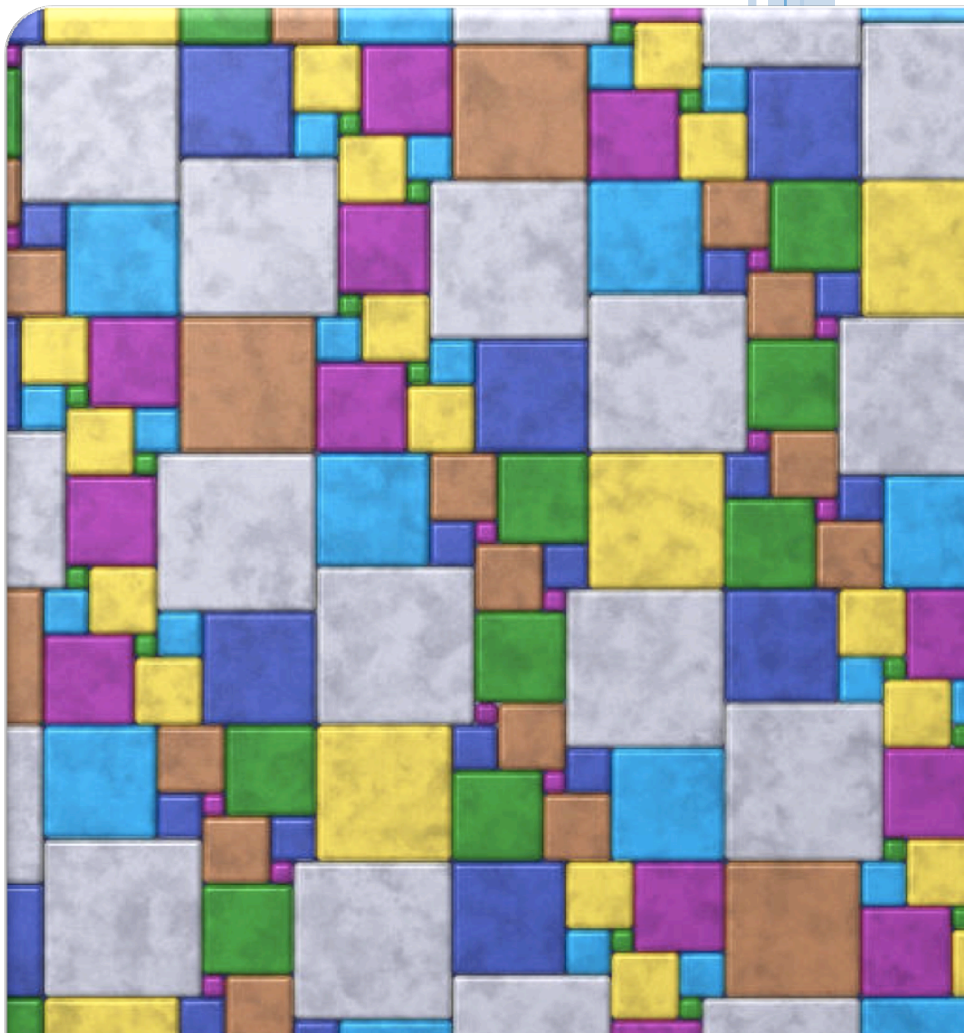
Scalability

Frequency

Comparability
Can give better/worse judgments

Flexibility
Can apply to different types and contexts





Comparability
Can give better/worse judgments

Flexibility
Can apply to different types and contexts

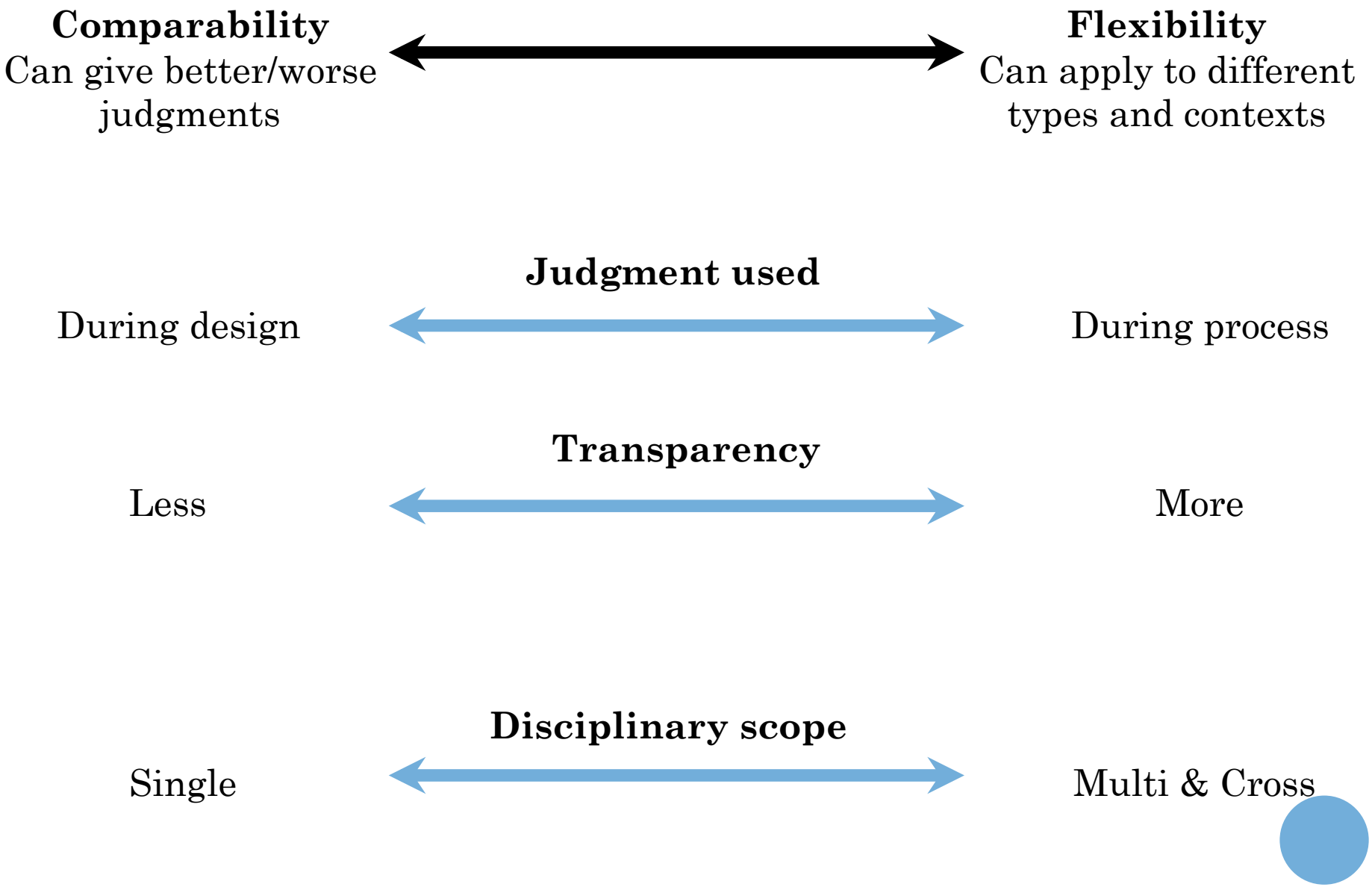
Collect a few, the same for each
eg papers, PhDs

Collect lots, the same for each
eg papers, PhDs, Patents, guideline citations

Collect lots, different for each

Comparability between units

Comparability over time


Comparability
Can give better/worse judgments
Flexibility
Can apply to different types and contexts
Judgment used
During design
During process
Transparency
Less
More
Disciplinary scope
Single
Multi & Cross

Assessment  **Improvement**

(Advocacy, Allocation, Accountability)

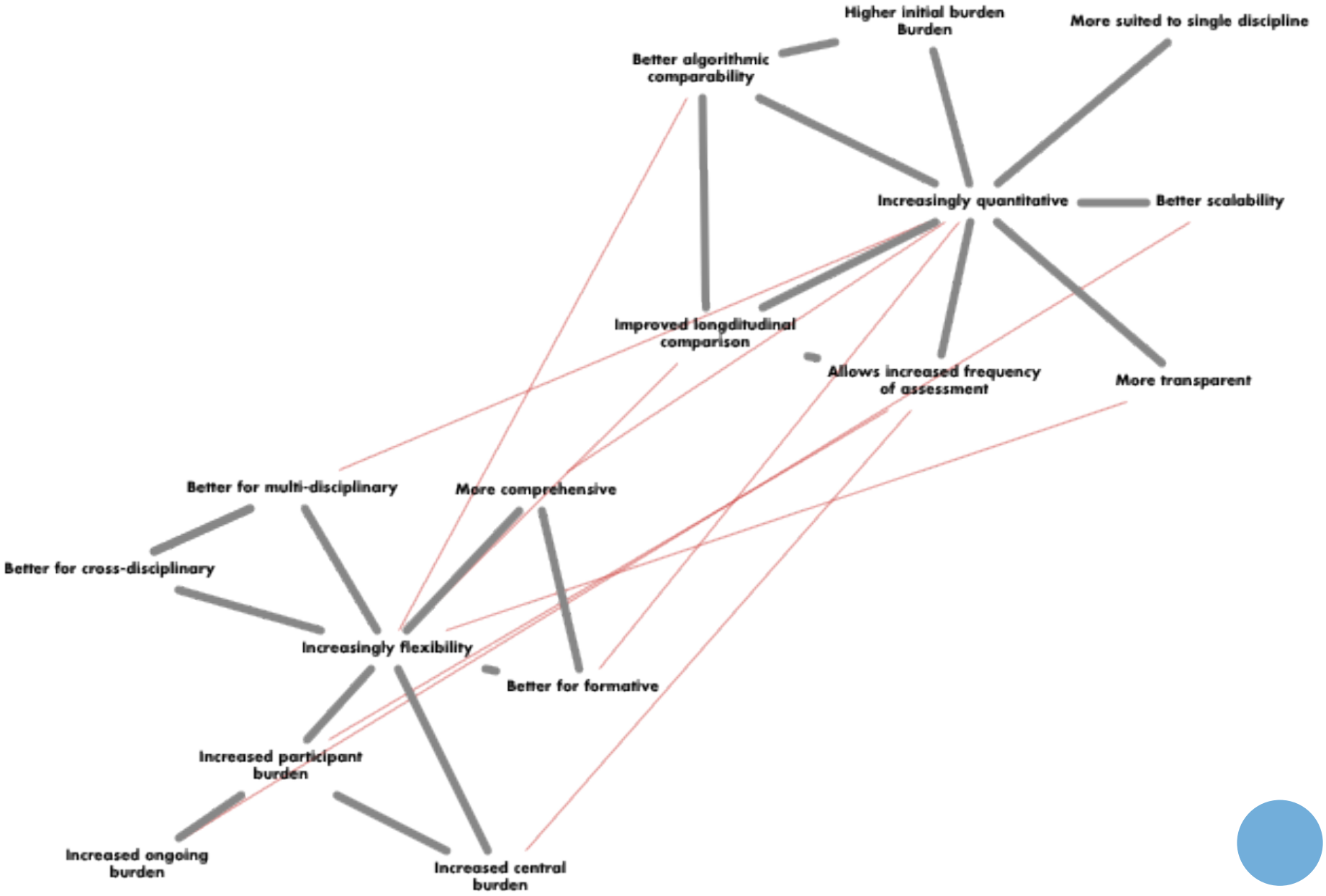
(Analysis)

Emphasis

Hard to game  Detail and understanding

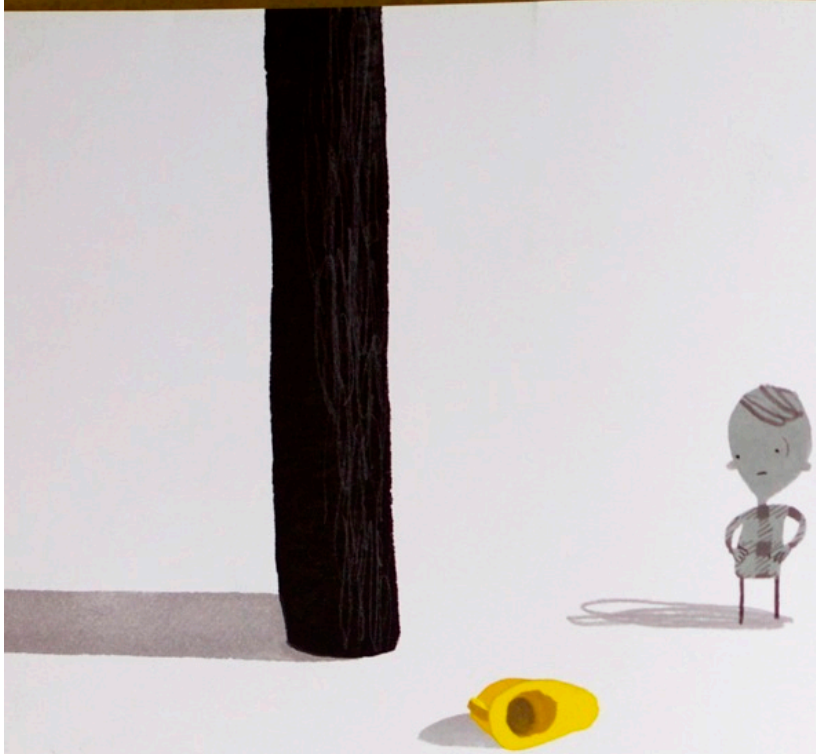
Transparency

More  Less


Higher initial burden
Burden
More suited to single discipline
Better algorithmic comparability
Increasingly quantitative
Better scalability
Improved longditudinal comparison
Allows increased frequency of assessment
More transparent
Better for multi-disciplinary
More comprehensive
Better for cross-disciplinary
Increasingly flexibility
Better for formative
Increased participant burden
Increased ongoing burden
Increased central burden

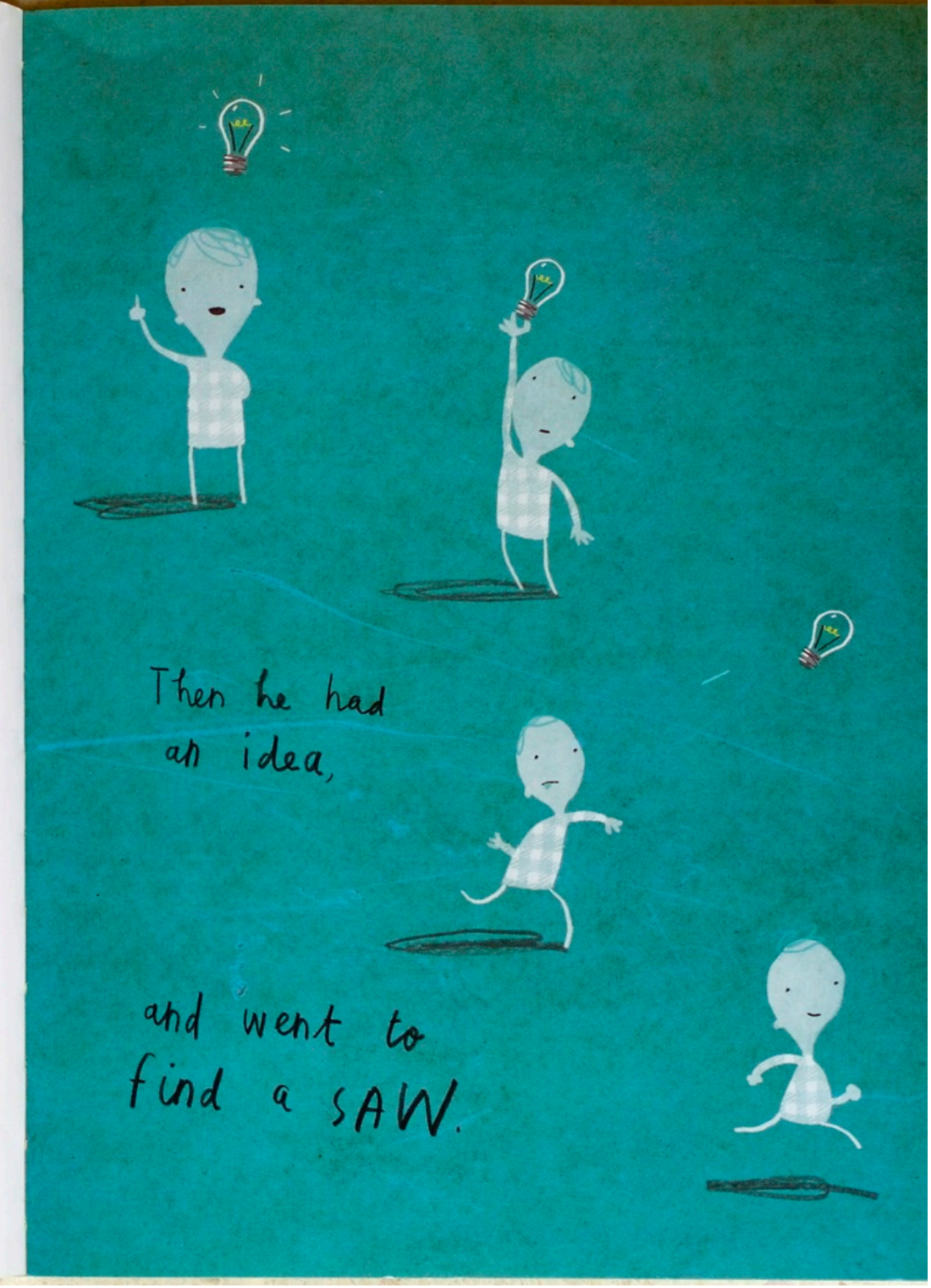
Recap/Summary

- Looked at how we know stuff
  - Theory based
  - Experimental/comparative
  - We need both
- Used logic models to explore how our programme works
- Introduced the Payback framework as exemplar framework
- Reviewed the diversity of frameworks
- Explored the tensions that underlie research impact assessment



And there they stayed,
stuck between the orang-utan
and one of the BOATS.

Firemen would DEFINITELY
be noticed missing and
Floyd KNEW he'd be in
BIG TROUBLE!

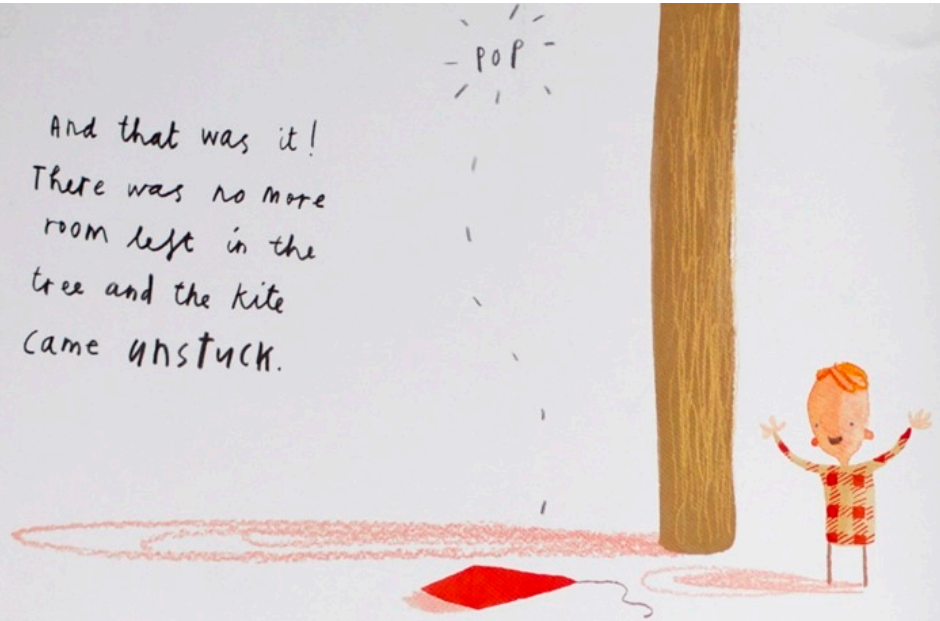


Then he had
an idea,

and went to
find a SAW.

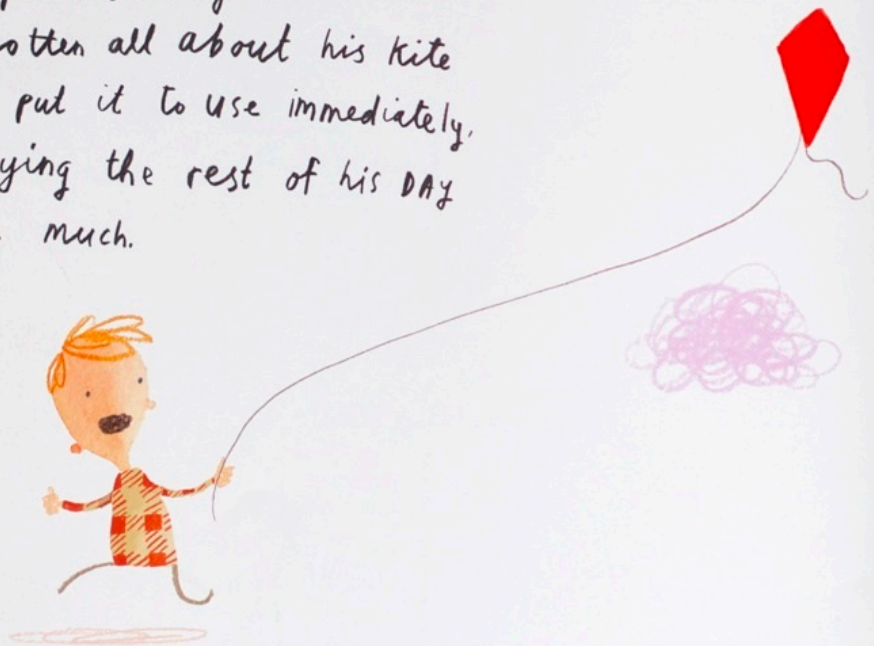
He lined it
up as best he could...

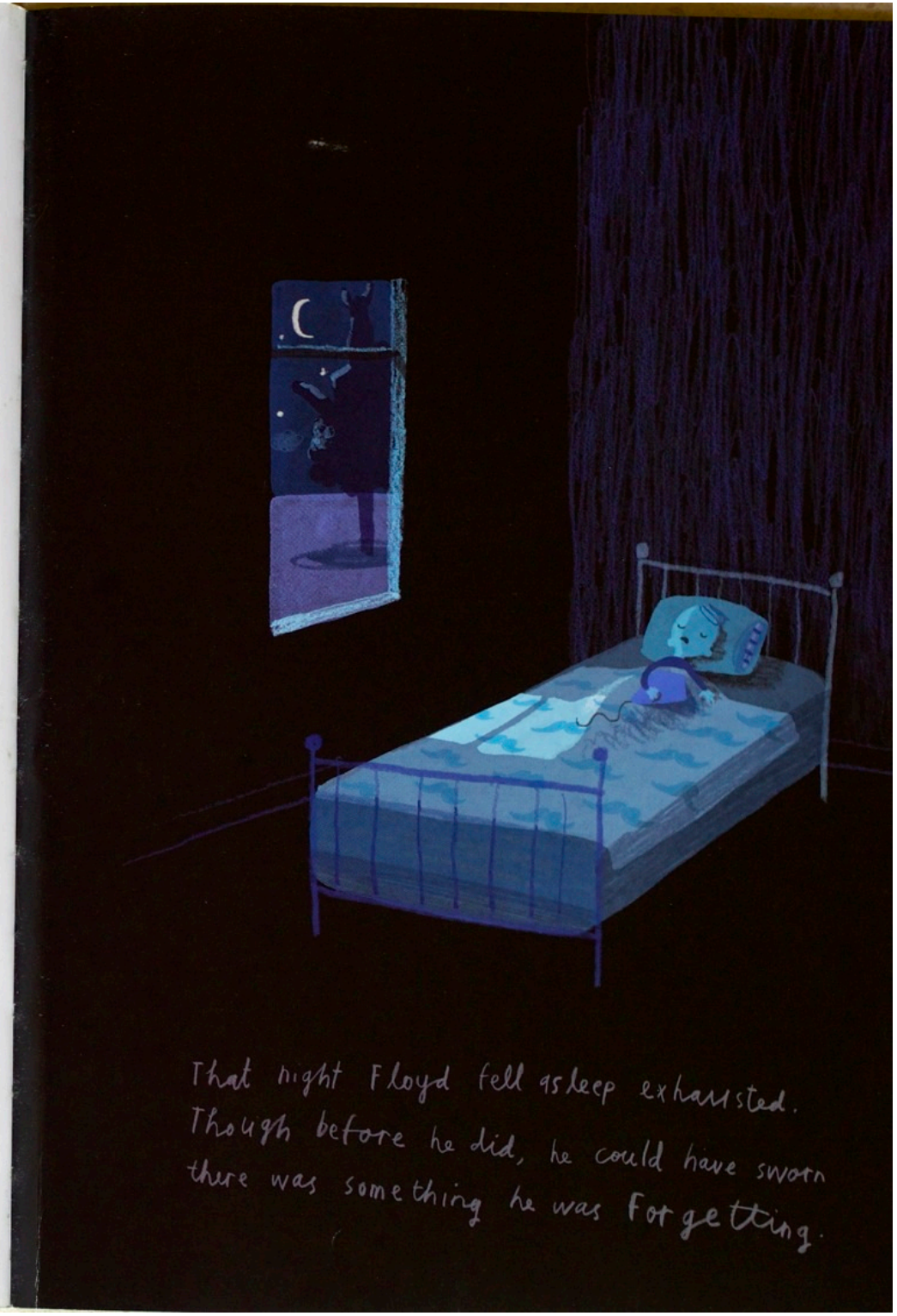
...and HURLED IT UP THE TREE.



And that was it! There was no more room left in the tree and the kite came unstuck.

Floyd was delighted. He had forgotten all about his kite and put it to use immediately, enjoying the rest of his day very much.





That night Floyd fell asleep exhausted. Though before he did, he could have sworn there was something he was forgetting.

Questions?

Steven Wooding
RAND Europe
@drstevenwooding

RAND EUROPE

Recommended Reading 1

- W.K. Kellogg Foundation Logic Model Development Guide
  - http://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide
- University of Wisconsin – Extension, Division of Cooperative Extension. Program Development and Evaluation Unit (PDE). http://www.uwex.edu/ces/pdande/evaluation/evallogicmodel.html
- Jordan, Gretchen B. 2013. Logic Modeling: A Tool for Designing Program Evaluations, in Handbook on the Theory and Practice of Program Evaluation, Albert N. Link and Nicholas S. Vonortas, Editors, Edward Elgar Publishing.
- Funnell, S. (2000). "Developing and Using a Program Theory Matrix for Program Evaluation and Performance Monitoring," in New Directions for Evaluation, Rogers, et.al. Eds., San Francisco: Jossey-Bass, Number 87, Fall, pp. 91-102.
- Jeffers, O. (2011). "Stuck", HarperCollins, London