

ICSTI 2014 Annual Conference in Tokyo  
20-21 October 2014

# **Towards indicators for ‘opening up’ science and technology policy**

Ismael Rafols, Tommaso Ciarli and Andy Stirling

*Ingenio* (CSIC-UPV), Universitat Politècnica de València &  
SPRU —Science and Technology Policy Research, University of Sussex

# Summary of the argument

- From information management (1st day)  
....to **using information for managing science** (2nd day)
- **Great potential** of new information infrastructure to provide new **insights for science policy**
  - **Many traces of research dimensions so far hidden dimensions**
    - Faster capture of scientific impact (downloads, Mendeley,)
    - Interactions with social actors (press release, twitter, etc.)
    - Activities previous not accounted (data sharing)
- **However...** need to foster a wise use of information data
  - **All Techs have intended and unintended effects**
  - **By focusing attention in some types of measures, bias against others**
    - Streetlight effect and drunkard's search.
- **Proposal:** it is not only about MORE indicators. It is about what is the QUALITIES of indicators. Putting questions to foster judgement (opening up), rather than reducing options (closing down).

1. Why are we mapping science?

The role of measurement in science advice

# On the role of expert advice in policy

(researchers on science dynamics provide expertise for science policy)

## The linearity-autonomy model of expert advice (Jasanoff, 2011)

- **Expert** knowledge is the **best possible foundation** for public decisions
- Experts should establish the facts that matter **independently**.
  - S&T indicators produce evidence of these facts.











## However, this model has been challenged

- “... society or the public sphere can, in principle, provide a *more rational* solution to political controversy than that offered by the application of technical methods.” (Barry, 2001, p. 8)
- Scientific trajectories are often shaped by pressures which are not always aligned with wider public good (Roger and Pielke, 2007)

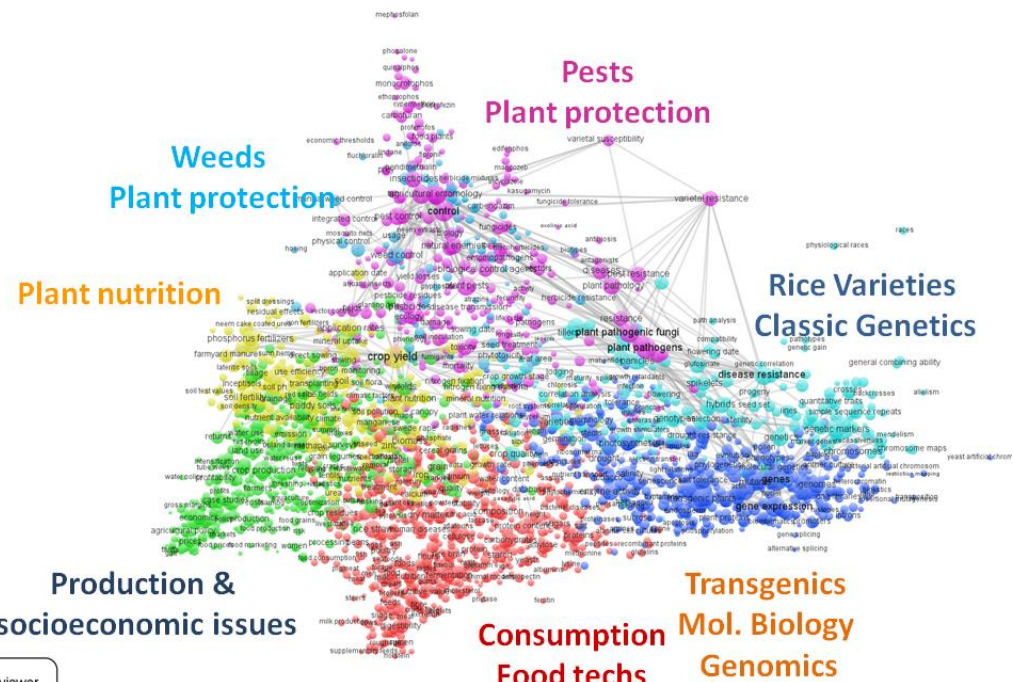
**What is (should be) the role of STI indicators in policy advice?**

# What type of “answer” should advice provide? How can S&T indicators help in science policy?

**Model 1: proposing “best choices”**  
Rankings -- ranking list of preferences

Academic Ranking of World Universities - 2011				
World Rank	Institution	Country	National Rank	Total Score
1	Harvard University		1	100.0
2	Stanford University		2	
3	Massachusetts Institute of Technology (MIT)		3	
4	University of California, Berkeley		4	
5	University of Cambridge		1	
6	California Institute of Technology		5	
7	Princeton University		6	
8	Columbia University		7	
9	University of Chicago		8	
10	University of Oxford		2	

**Model 2: exploring complementary choices**  
Research portfolios on rice



# Bad bibliometrics: The case of journal impact factor

- Journal Impact Factor (JIF), developed for **assessing journals**.
- Begins to be used to assess individual papers and researchers.
- In the 1990s, H. Moed and T. van Leeuwen → technical **inconsistencies**.  
Per Seglen → **inadequate for research evaluation** of papers or individuals.
- However, the use of **JIF thrived** for the 1990s and 2000s.
- **Reversal of causality:** Initially, the metrics reflected that reputation of a journal which reflected the reputation of the researchers involved in it. With JIS, the relationship was inverted: the metrics gave reputation to the journal, which gave reputation to the authors.
- San Francisco DORA (2013) heavily **criticised JIF –with strong political backing**. This led to media attention and perhaps decline on use of JIF.
- It is unclear that more accurate journal metrics (e.g. SNIP) are being widely used.
- Yet journal metrics for research assessment **remains a very common practice**.

# Morals of the JIF story

- **Indicators take a life of their own** and become used in contexts that are often inappropriate. JIF was not meant to be used in assessment. More appropriate metrics are not necessarily those adopted.
- **Indicators are performative**, i.e. they have an effect on who is measured since they signal what is perceived as "good performance" rather than just measuring "performance". This is why managers like them. Researchers change their behaviour.
- **Scientific "truth" does not win the debate** without a social context that supports it. Best advice on JIF was ignored for more than 20 years

Looking back at the problems of bibliometrics  
–will new metrics better?

## **Use of conventional STI indicators has been \*problematic\***

- Narrow inputs (pubs, pats...)
- Scalar outputs (rankings!) – misplaced concreteness.
- Aggregated solutions --missing group variation, error estimates
- Opaque selections and classifications  
Privately owned databases.
- Large, leading STI groups embedded in government / consultancy, with limited possibility public scrutiny



## From S&T indicators for justification and pushing...

### **Justification** in decision-making

- Weak justification, “Give me a number, any number!”
- Strong justification, “Show in numbers that X is the best choice!”

### S&T Indicators have a **performative** role:

- They don't just measure. Not ‘just happen to be used’ in science policy (neutral)
- They signal to stakeholders what is important.
- Articulate framings on what is good performance:
  - More pubs? More pats? Collaboration? Interdisciplinarity? Press releases?

## ... towards S&T indicators as tools for strategic thinking and deliberation

Yet is possible to design indicators that foster reflection rather than justifying or reinforcing dominant perspectives  
(this leads to decrease in diversity, driving down opportunities)

This shift is facilitated by trends pushed by information techs and visualisation tools

- More inputs (pubs, pats, but also news, webs, etc.)
- Multidimensional outputs (interactive maps)
- Multiple solutions -- highlighting variation, confidence intervals
- More inclusive and contrasting classifications (by-passing private data ownership? Pubmed, Arxiv)
- More possibilities for open scrutiny (multiple research groups)

## 2. Conceptual framework:

“broadening out” vs. “opening up” policy appraisal

# Policy use of S&T indicators: Appraisal

## **Appraisal:**

**‘the ensemble of processes through which knowledges are gathered and produced in order to inform decision-making and wider institutional commitments’ Leach et al. (2008)**

**Breadth:** extent to which appraisal covers diverse dimensions of knowledge

**Openness:** degree to which outputs provide an array of options for policies.

# Policy use of S&T indicators: Appraisal

## Appraisal:

**‘the ensemble of processes through which knowledges are gathered and produced in order to inform decision-making and wider institutional commitments’ Leach et al. (2010)**

Example:

Allocation of resources based on research “excellence”

**Breadth:** extent to which appraisal covers diverse dimensions of knowledge

Narrow: citations/paper

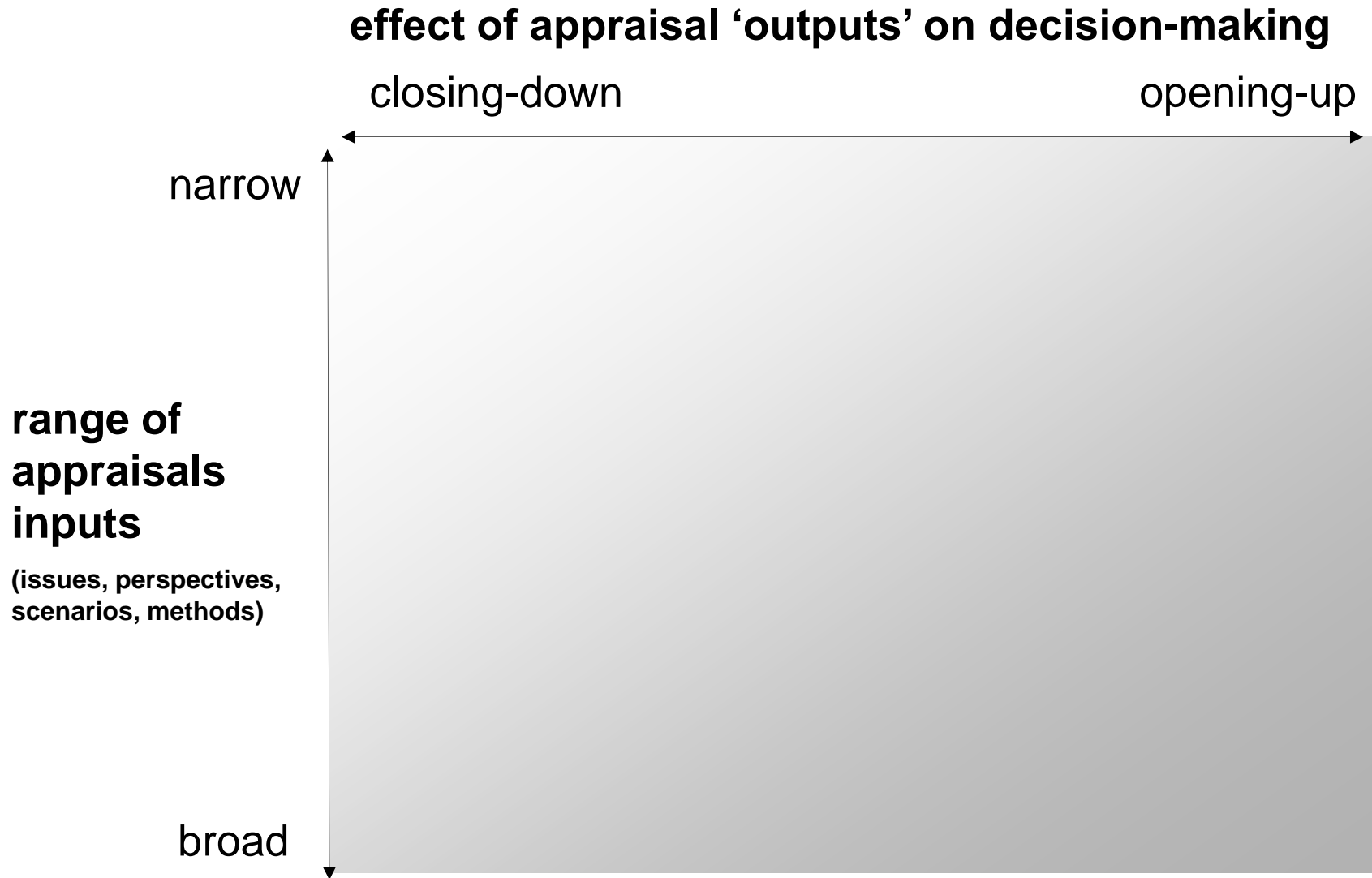
Broad: citations, peer interview, stakeholder view, media coverage, altmetrics

**Openness:** degree to which outputs provide an array of options for policies.

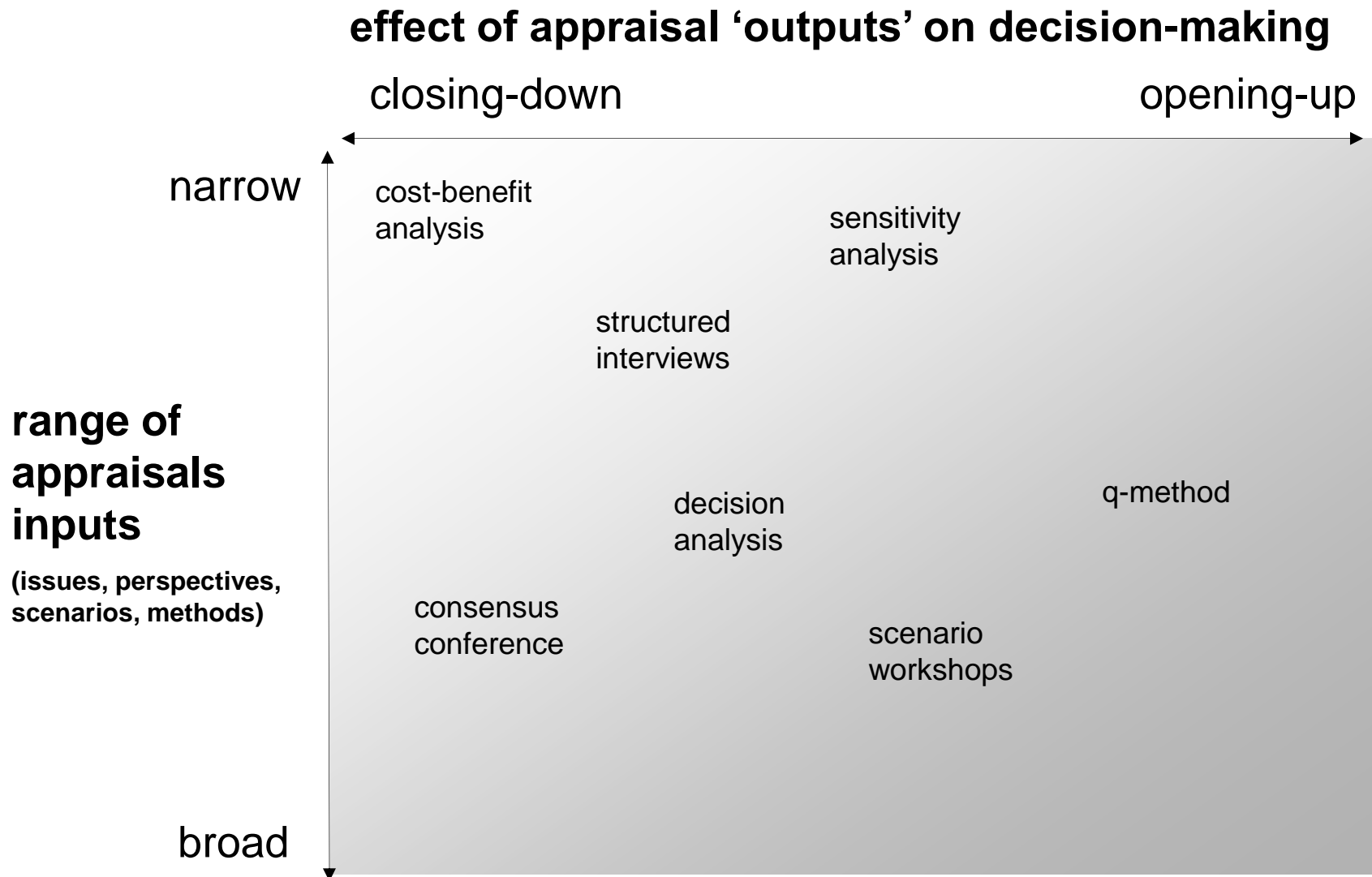
Closed: fixed composite measure of variables → unitary and prescriptive

Open: consideration of various dimensions → plural and conditional

# Appraisal methods: broad vs. narrow & closing vs. opening



# Appraisal methods: broad vs. narrow & close vs. open

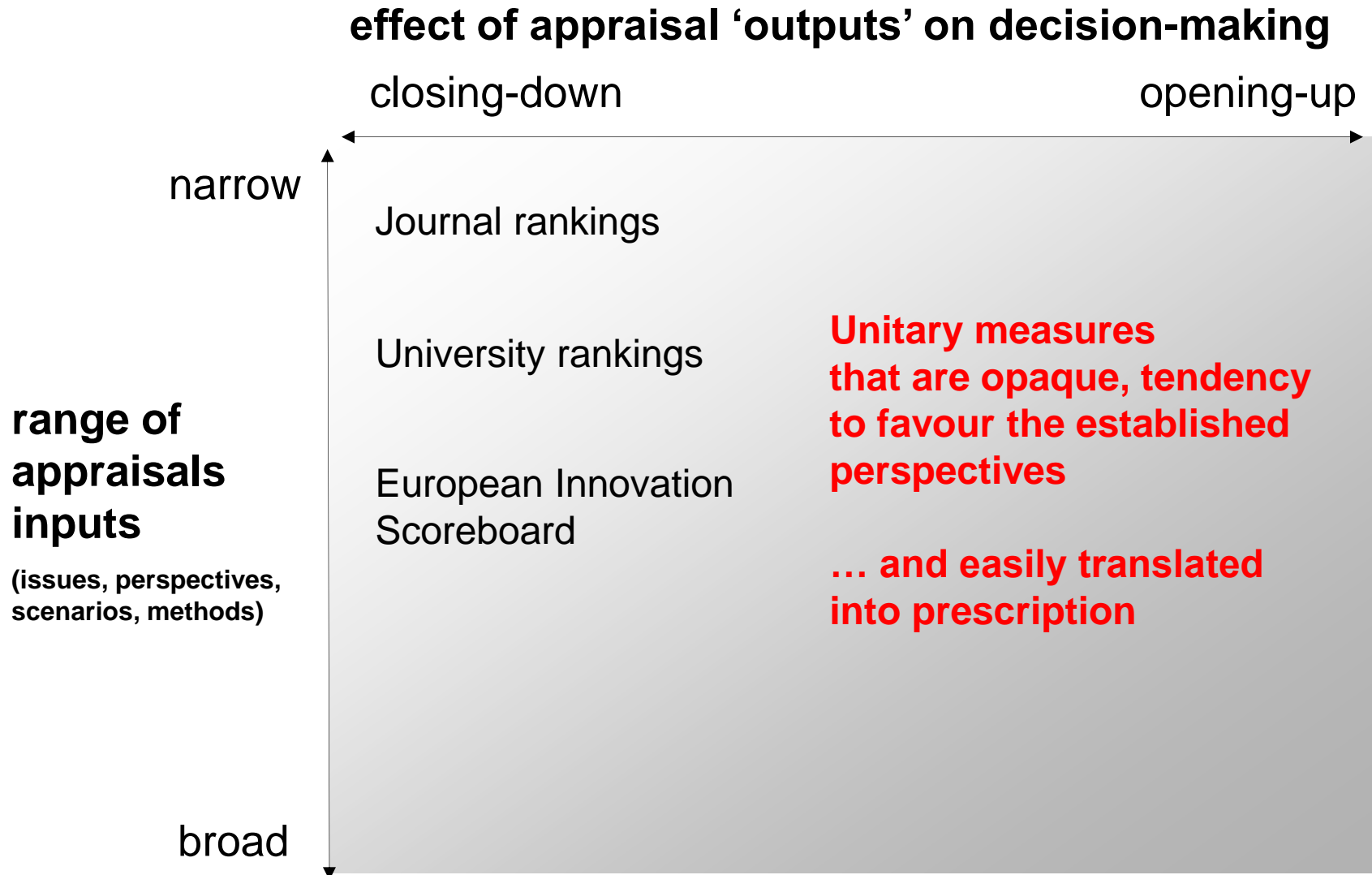






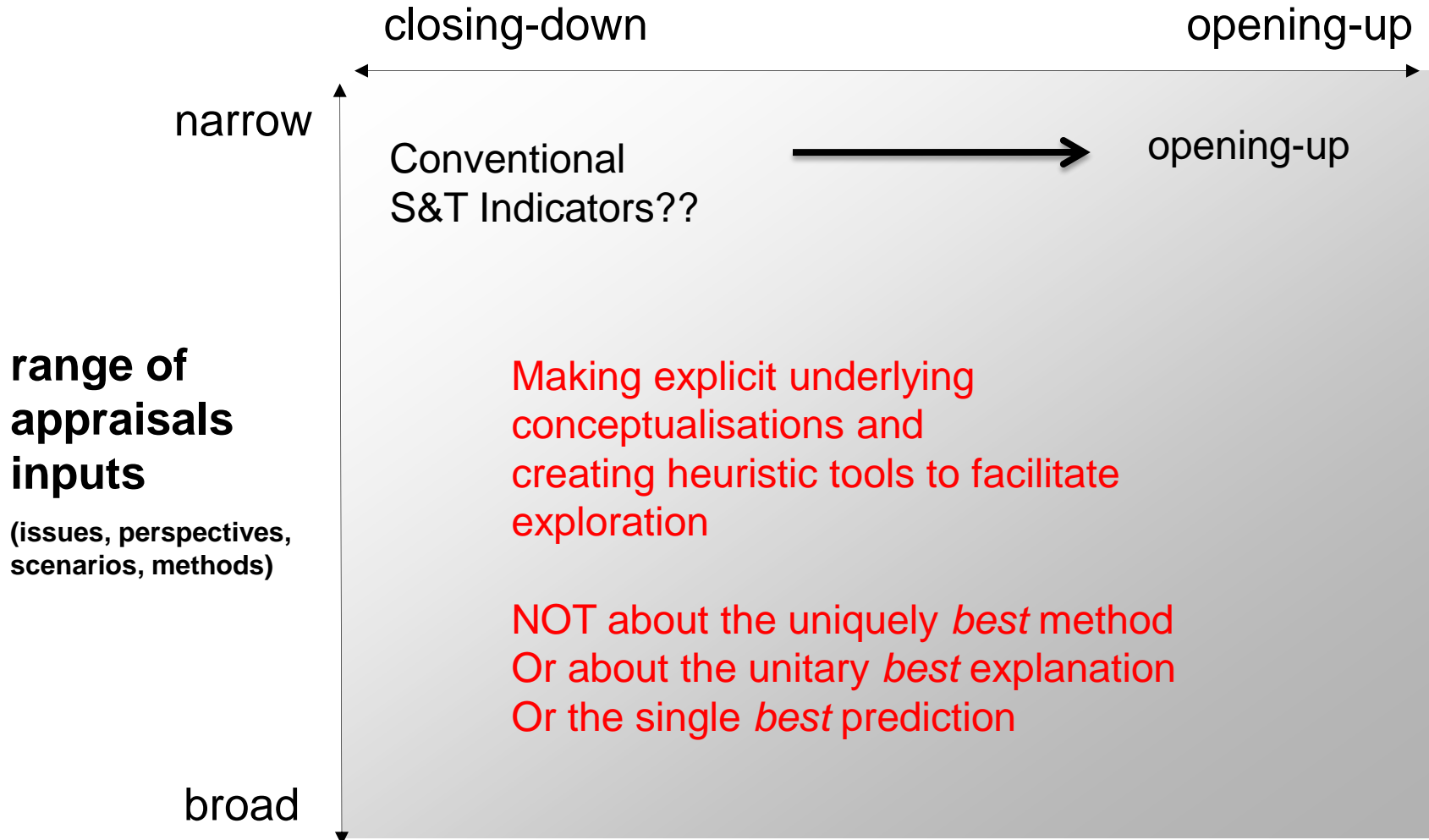


# Appraisal methods: broad vs. narrow & closing vs. opening



# Opening up S&T Indicators

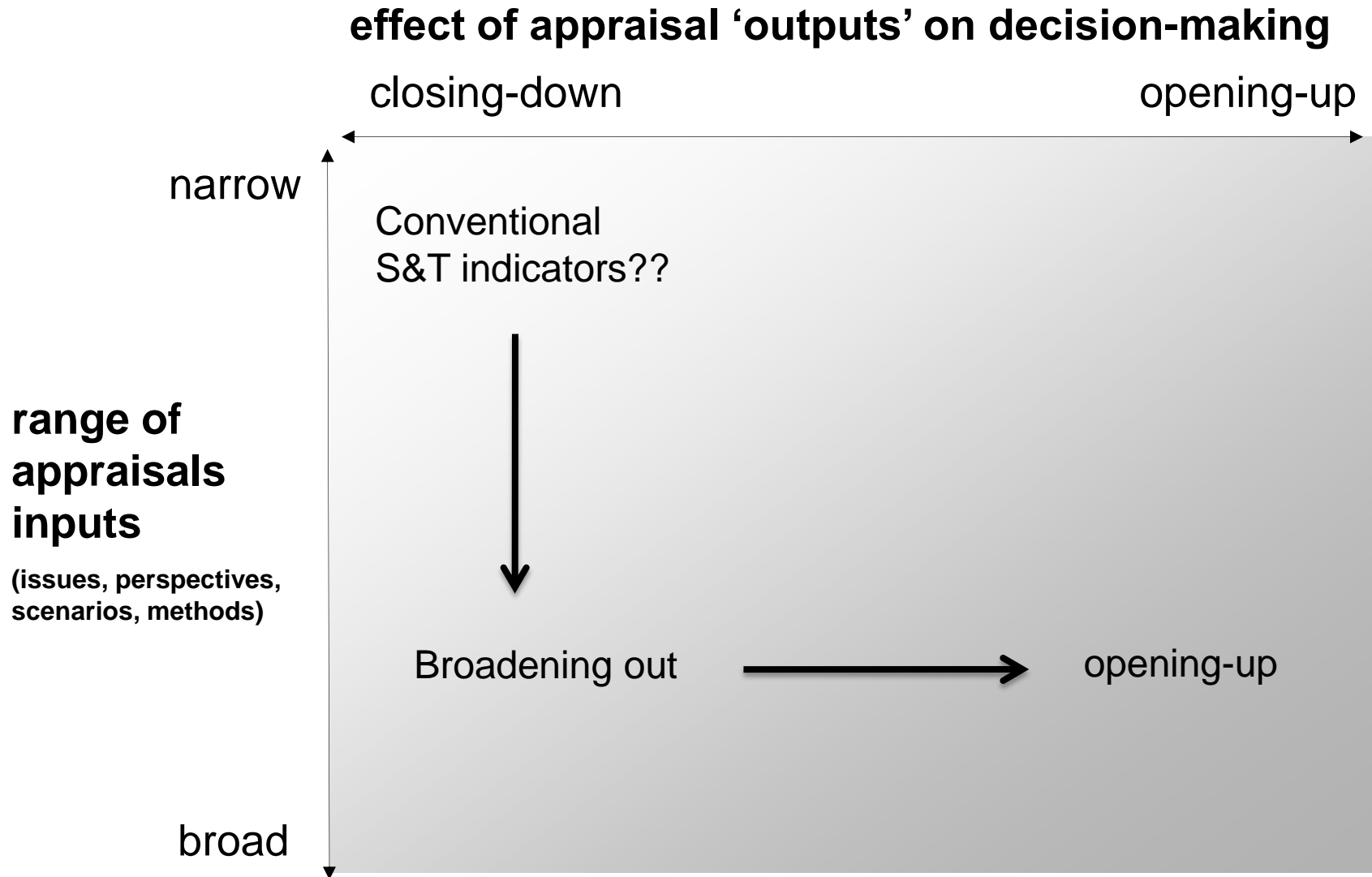
## effect of appraisal 'outputs' on decision-making



## 2. Examples of Opening Up

- a. Broadening out AND Opening up
- b. Opening up WITH NARROW inputs

# 1. Preserving multiple dimensions in broad appraisals



# Composite Innovation Indicators (25-30 indicators)

## European (Union) Innovation Scoreboard

Grupp and Schubert (2010) show that order is highly dependent on indicators weightings.

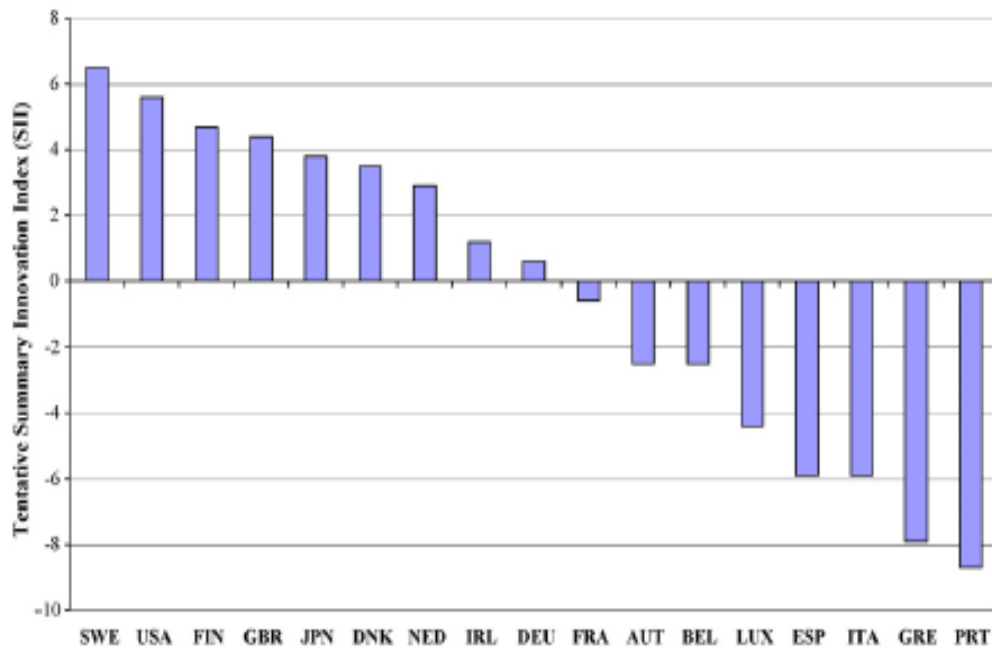


Fig. 1. Summary Innovation Index of the European Innovation Scoreboard 2001 (modified graph from European Commission, 2001, p. 12).

## Sensitivity analysis

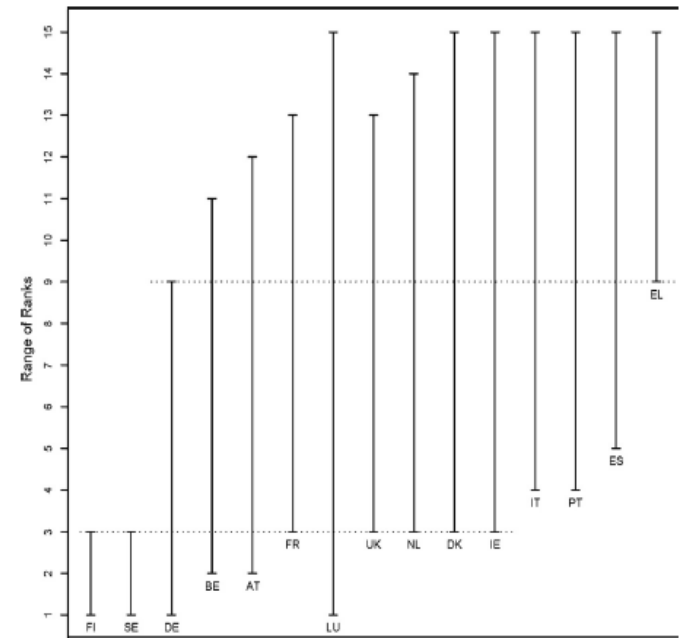
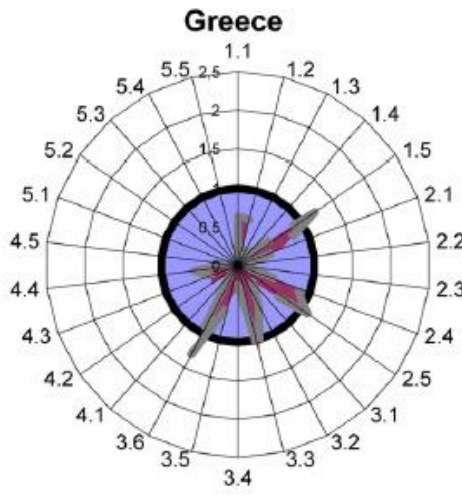
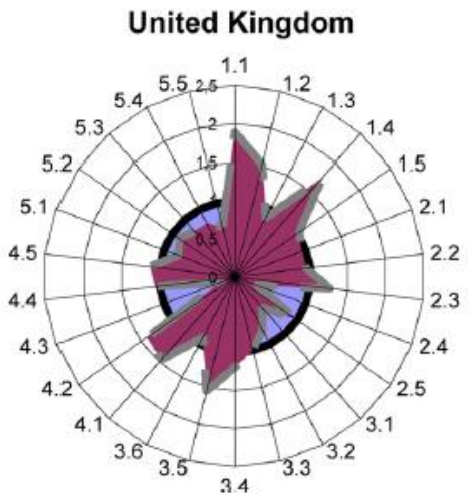
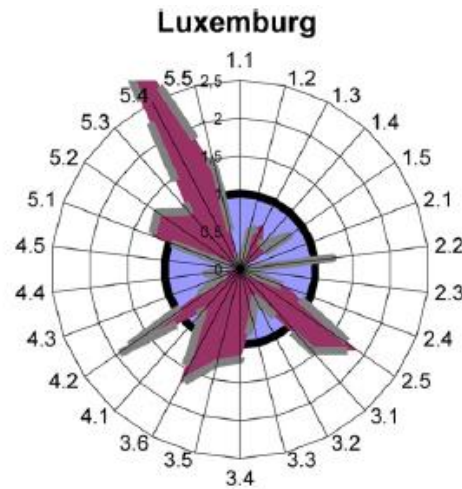
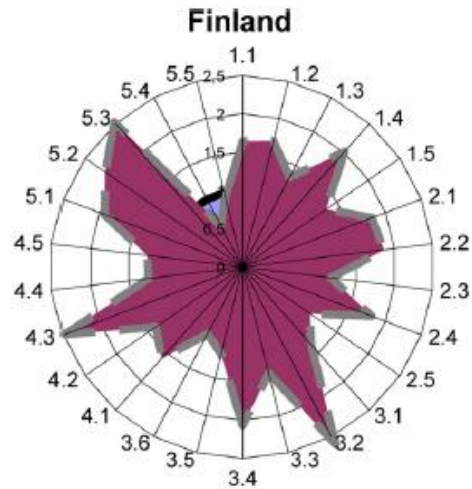


Fig. 2. Sensitivity analyses for the European Innovation Scoreboard 2005 (sources: original data European Commission, 2005).

# Solution: representing multiple dimensions

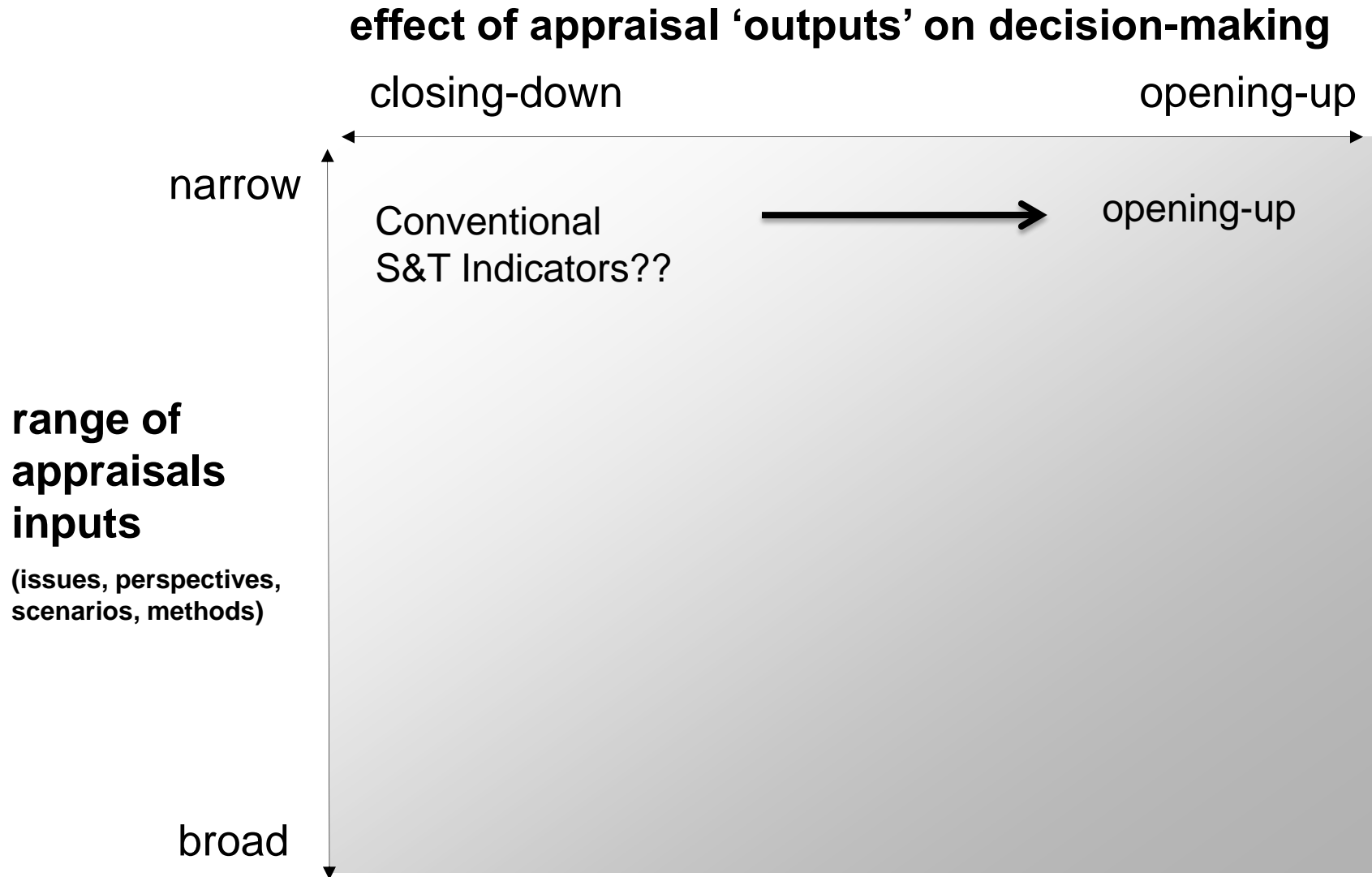
(critique by Grupp and Schubert, 2010)



Use of spider diagrams allows comparing like with like

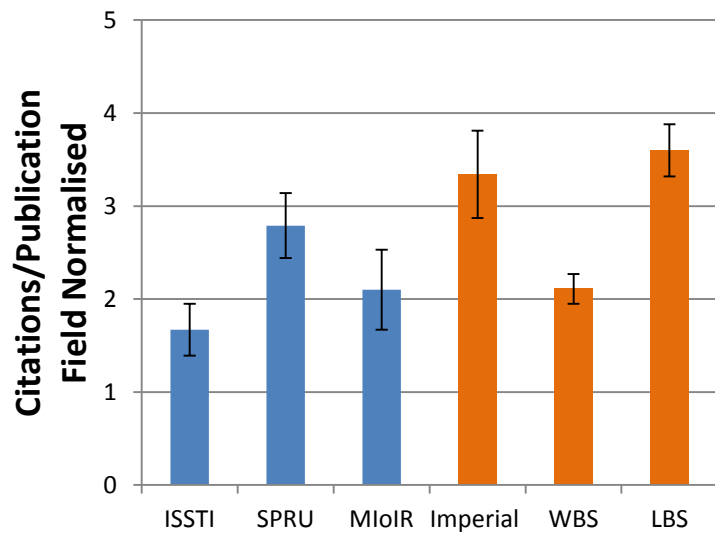
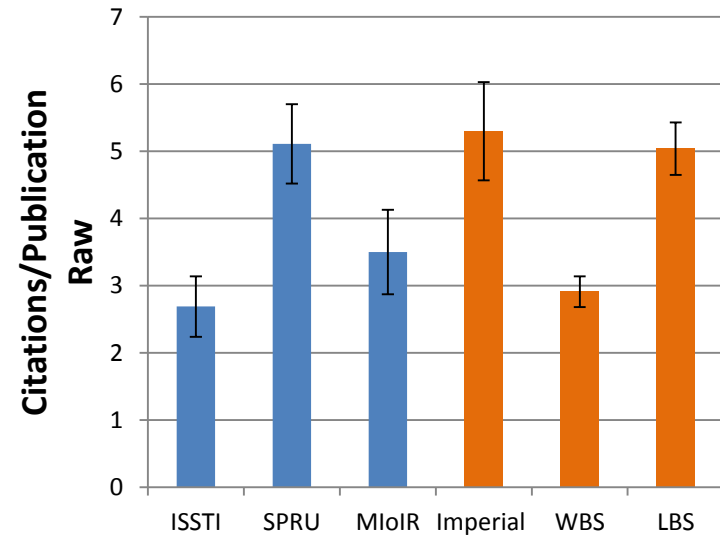
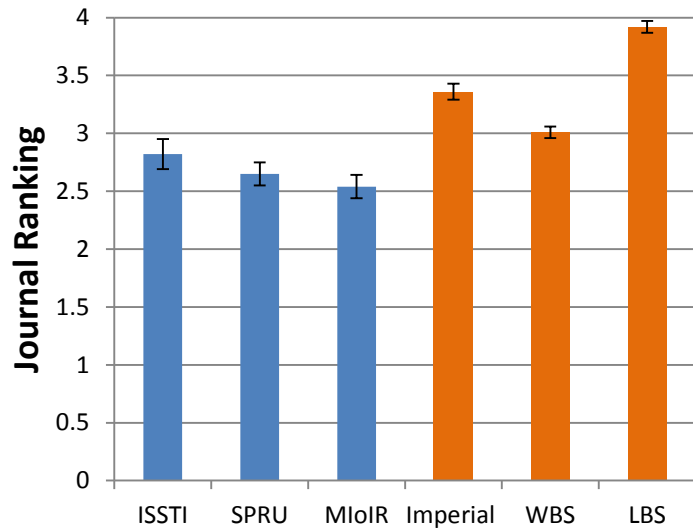
U-rank,  
University performance  
Comparison tools  
(Univ. Twente)

## 2. Opening up in spite of narrow inputs





# Citation performance, under different normalisations



# What is research “excellence”?

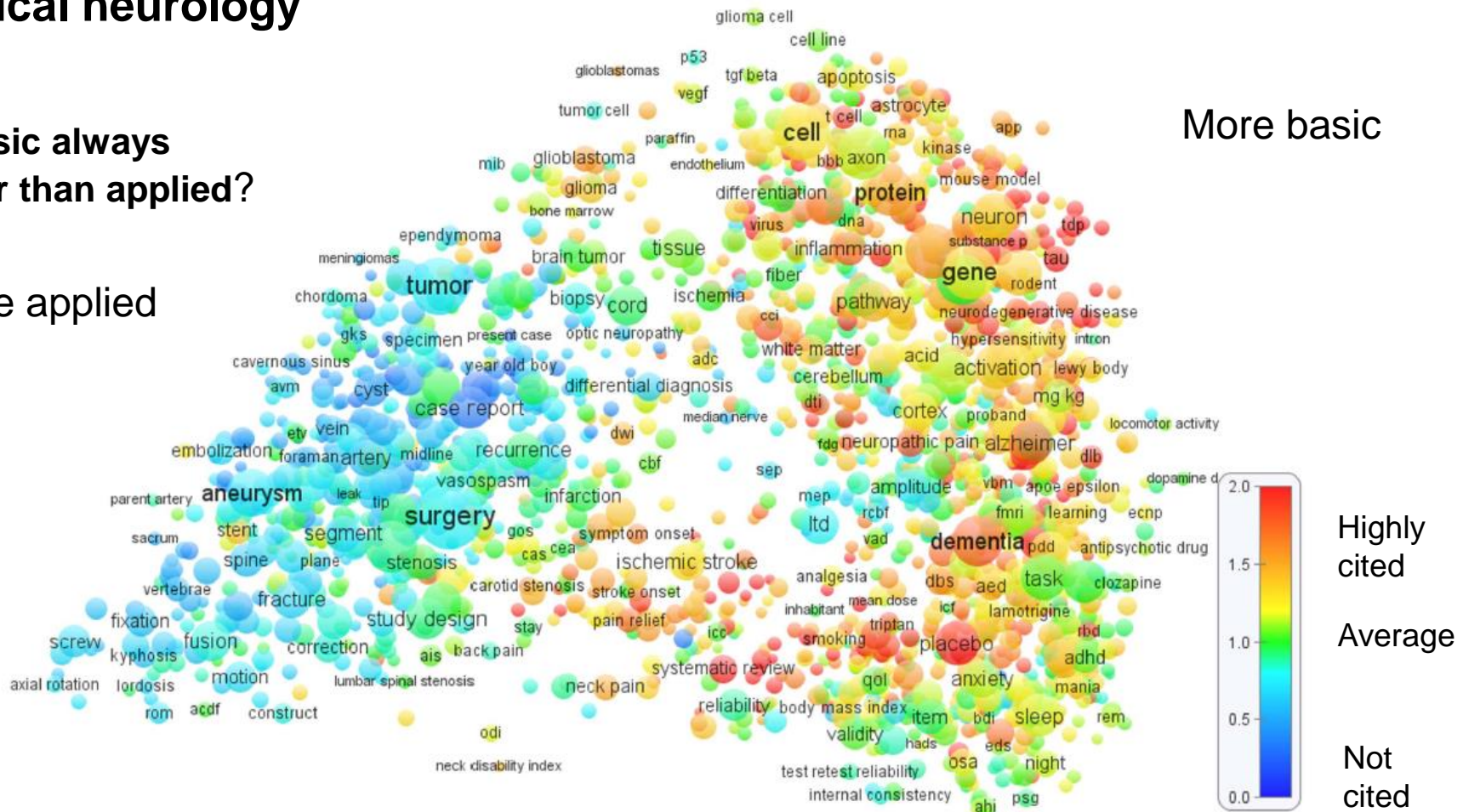
Citations: not stable to changes in classification and granularity (Zitt et al., 2005; Adams et al., 2008).

## Clinical neurology

Is basic always better than applied?

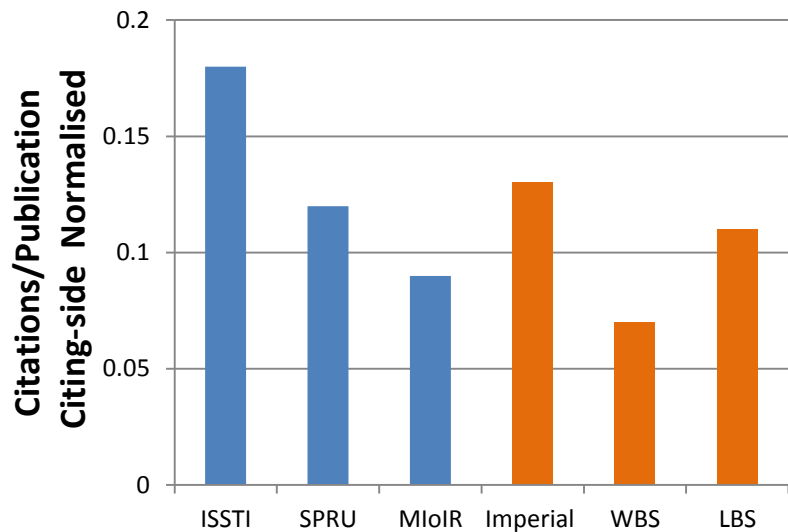
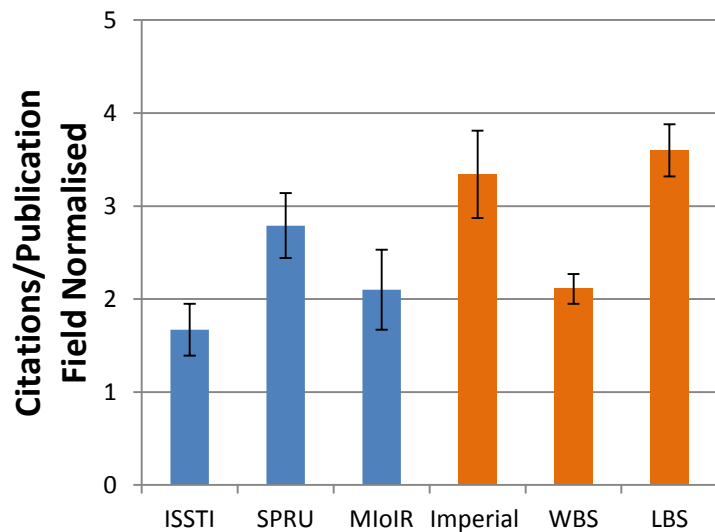
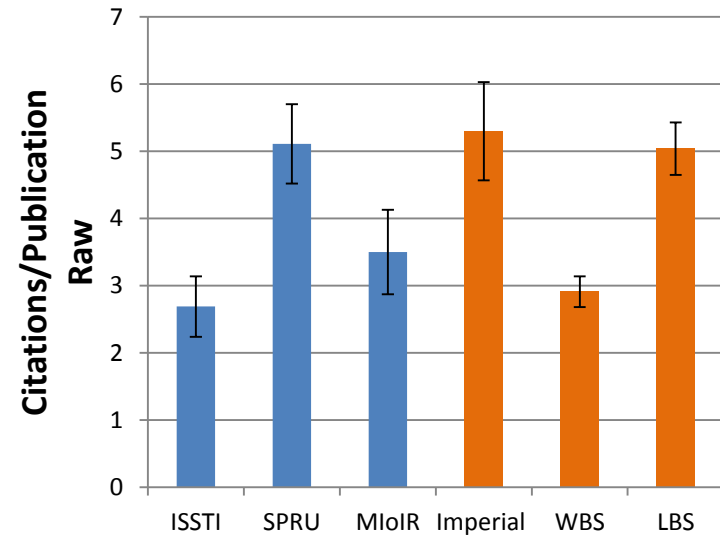
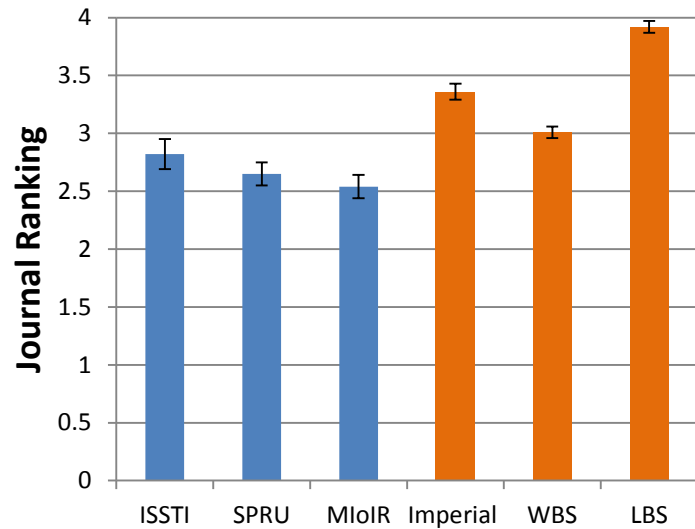
More applied

More basic



Van Eck, Waltman et al. (2013)

# Citation performance, under different normalisations



# Leiden ranking of Universities – includes sensitivity analysis

- Different measures of performance
  - Top 10%, mena number of pubs
- Under different conditions (language, fractional counting)
- Include confidence interval (bootstrapping)

Select indicators

Dimension of scientific performance:  ?













Rank universities based on:  ?  Show stability intervals ?

Select method of calculation

Normalize for university size ?

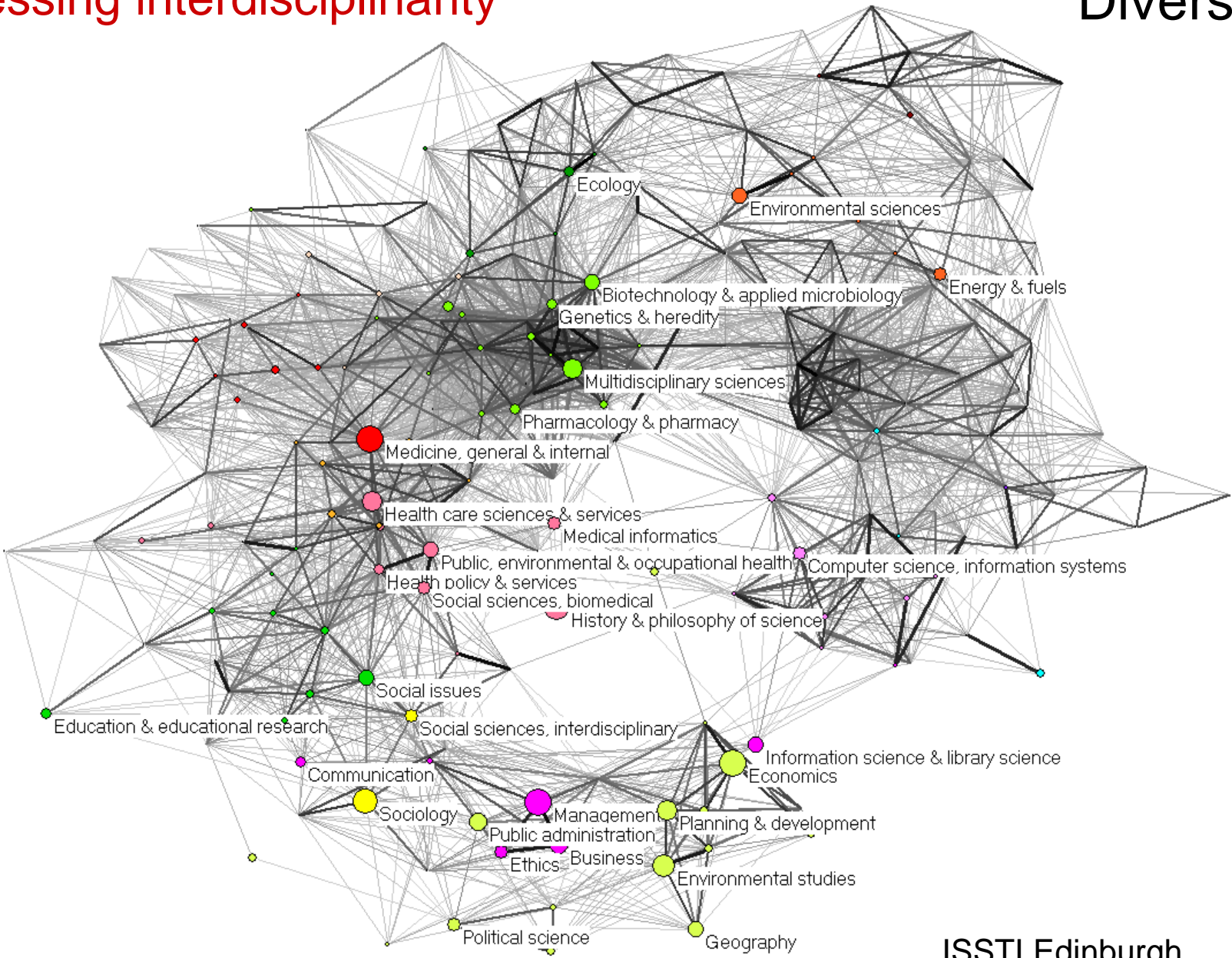
Assign collaborative publications fractionally to universities ?

Leave out non-English language publications ?

Rank	University	Country	P	PP <sub>top 10%</sub>	PP <sub>top 10%</sub> stability interval		
1	MIT		10465	25.2%			
2	Princeton Univ		5763	22.6%			
3	Harvard Univ		33511	22.5%			
4	Rice Univ		2635	22.2%			
5	Stanford Univ		15032	21.9%			
6	Caltech		6569	21.7%			

# Assessing interdisciplinarity

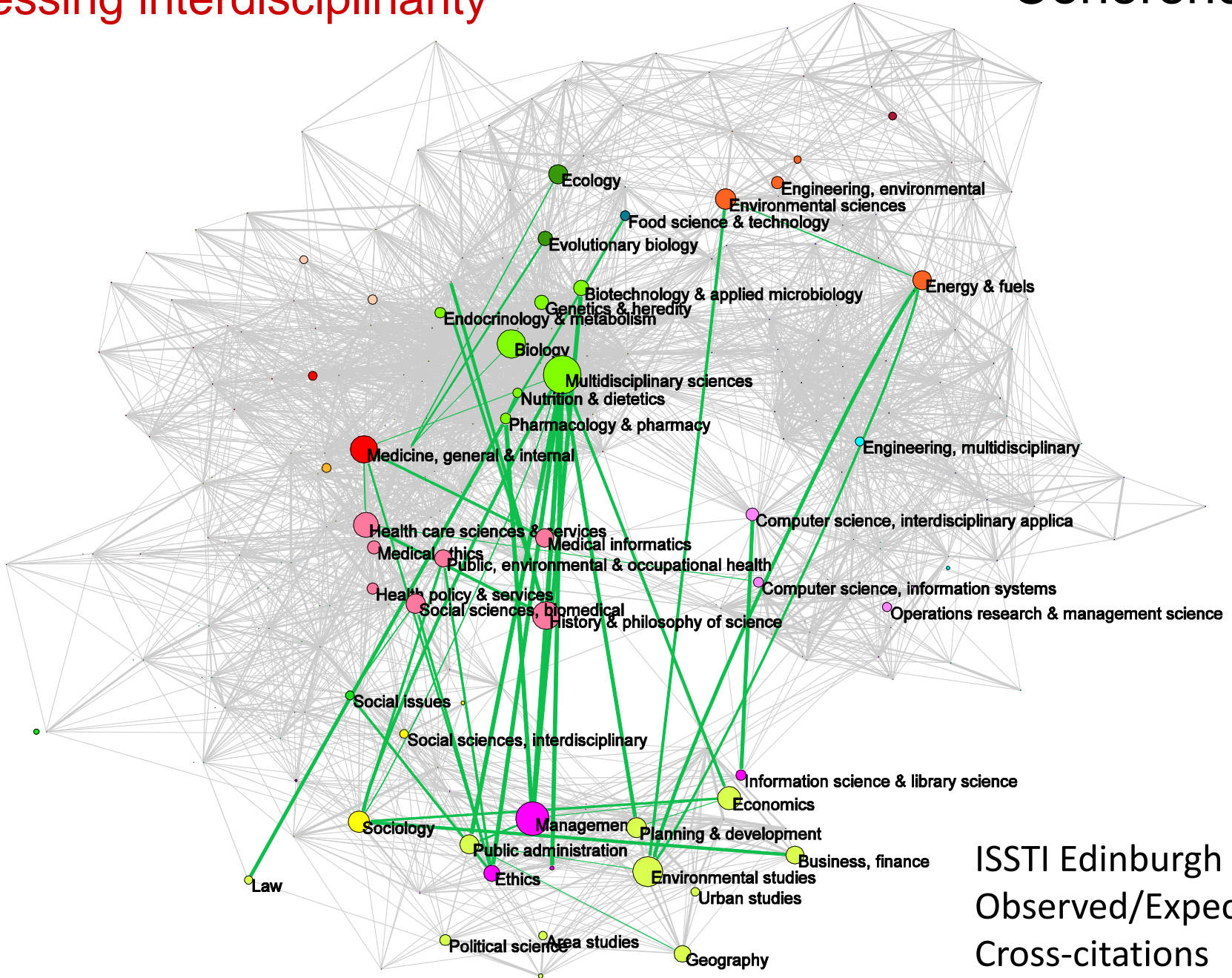
# Diversity





# Assessing interdisciplinarity

# Coherence

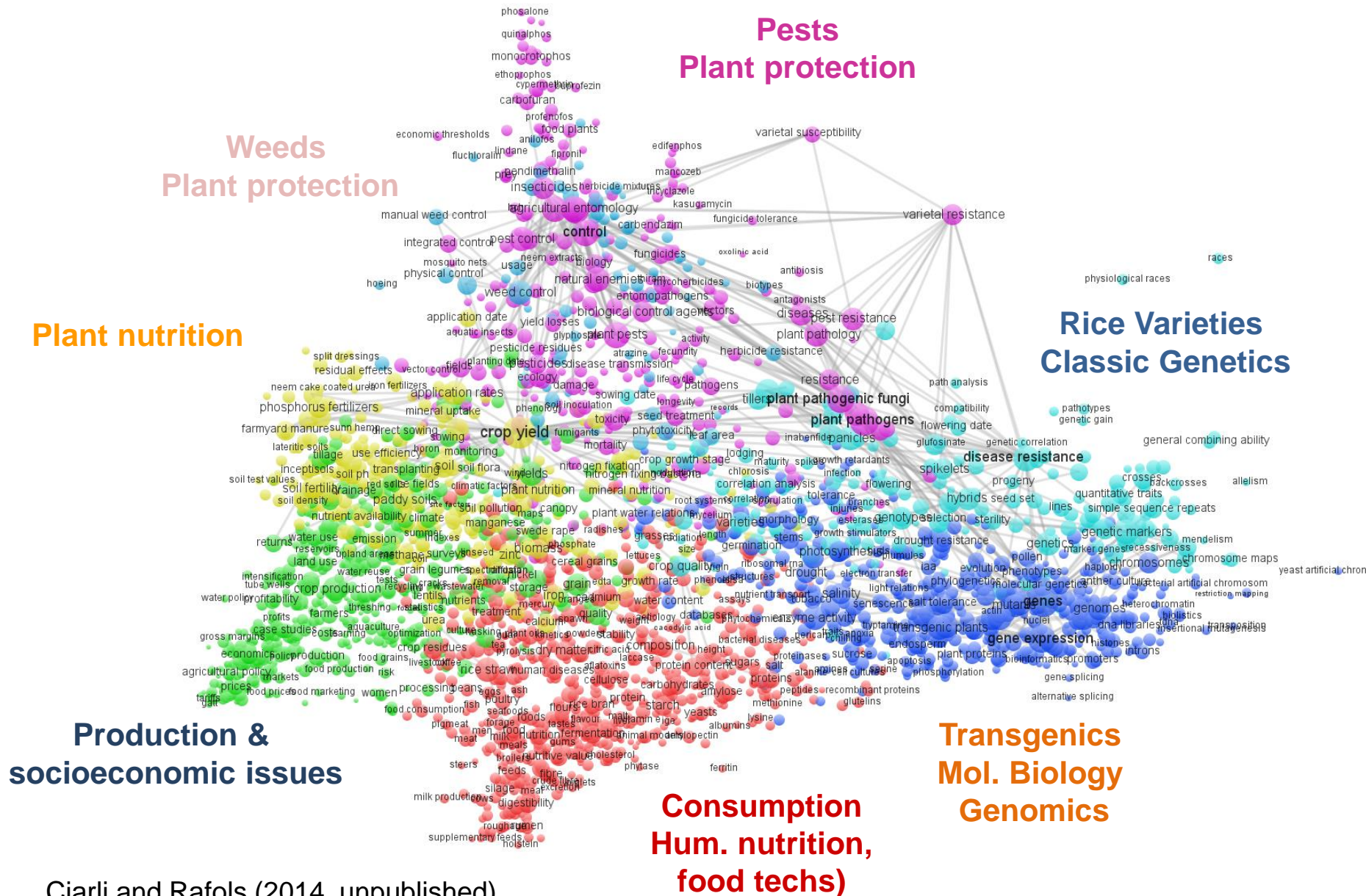


ISSTI Edinburgh  
Observed/Expected  
Cross-citations





# Thinking in terms of research portfolios: the case of rice



Ciarli and Rafols (2014, unpublished)









## 3. Conclusions

# S&T indicator as a tools to open up the debate

- ‘Conventional’ use of indicators (‘Pure scientist’ --Pielke)
  - Purely analytical character (i.e. free of normative assumptions)
  - Instruments of objectification of dominant perspectives
  - Aimed at legitimising /justifying decisions (e.g. excellence)
  - Unitary and prescriptive advice
- Indicators for Opening up (‘Honest broker’ --Pielke)
  - Aimed at locating the actors in their context and dynamics
    - Not predictive, or explanatory, but exploratory
  - Construction of indicators is based on choice of perspectives
    - Make explicit the possible choices on what matters
  - Supporting debate
    - Making science policy more ‘socially robust’
  - Plural and conditional advice

## Strategies for opening up or how to “keep it complex” yet “manageable”

- Presenting contrasting perspectives or “qualities”
  - At least TWO, in order to give a taste of choice
- Simultaneous visualisation of multiple properties / dimensions
  - Allowing the user take its own perspective
- Interactivity
  - Allowing the user give its own weigh to criteria / factors
  - Allowing the user manipulate visuals

This reflection --- Making explicit a trend that is already happening.

END OF PRESENTATION

# On the role of scientific advice in policy

The linearity-autonomy model of scientific advice (Jasanoff, 2011)

- **Scientific** knowledge is the **best possible foundation** for public decisions
- Scientists should establish **independently** the facts that matter.

The model has been adopted in science management,

**With STI indicators as evidence of the facts that matter.**

precisely after being heavily challenged (Pielke, 2007)

The debate is part of:

**What is the role of STI indicators in policy advice?**

(Building on the use of indicators in policy) (*Stirling, van Zwanenberg*)



# Do conventional indicators tend to favour incumbents?

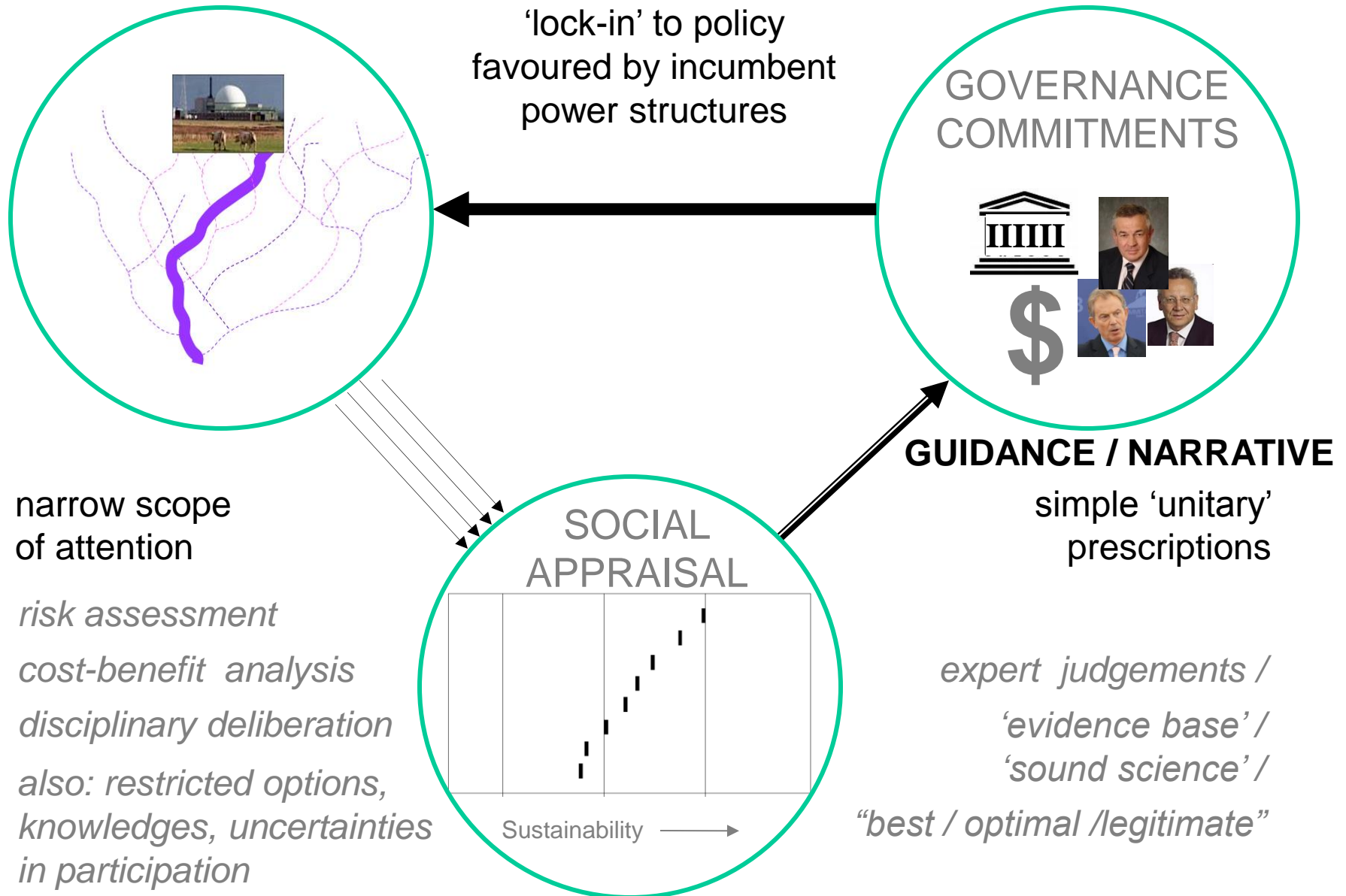
Hypothesis:

Elites and incumbents (directly or not) influence choice of indicators, which tend to benefit them... (Arnold, today)

“knowledge enables power, but power structures knowledge”  
(Stirling, 2012)

- **Crown indicator –Standard measure of performance (~1990-2010)**
  - ‘systematic underrating of low-ranked scientists’ (Opthof and Leydesdorff, 2010) (Not spotted for 15 years!)
- **Journal rankings in Business and Management.**
  - systematic underrating of interdisciplinary (heterodox) depts. (Rafols et al., 2012).
- **Others?? H-index**

# Conventional Policy Dynamics



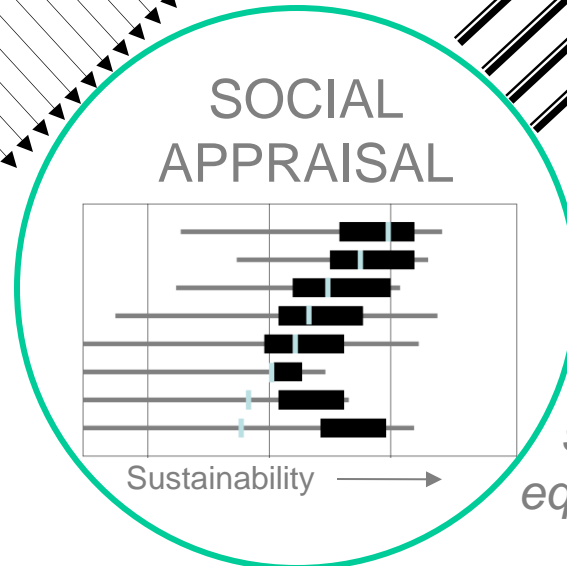
# Breadth, Plurality and Diversity

dynamic portfolios  
pursuing diverse socio-  
technical trajectories



broad-based  
processes of  
'precautionary appraisal'

*multiple: methods,  
criteria, options, frames,  
uncertainties, contexts,  
properties, perspectives*



'opening up' with  
'plural conditional'  
outputs to policymaking

*viable options under:  
conditions, dissonant views,  
sensitivities, scenarios, maps,  
equilibria, pathways, discourses*

