

Stepping Back to Progress Forwards



**Setting Standards for Meta-Evaluation
of Computational Creativity**

Anna Jordanous

Creativity Evaluation: methods



Graeme Ritchie (2001/2007) – Empirical Criteria



Simon Colton (2008) – Creative Tripod



Colton and Pease (2010-) – FACE/IDEA model

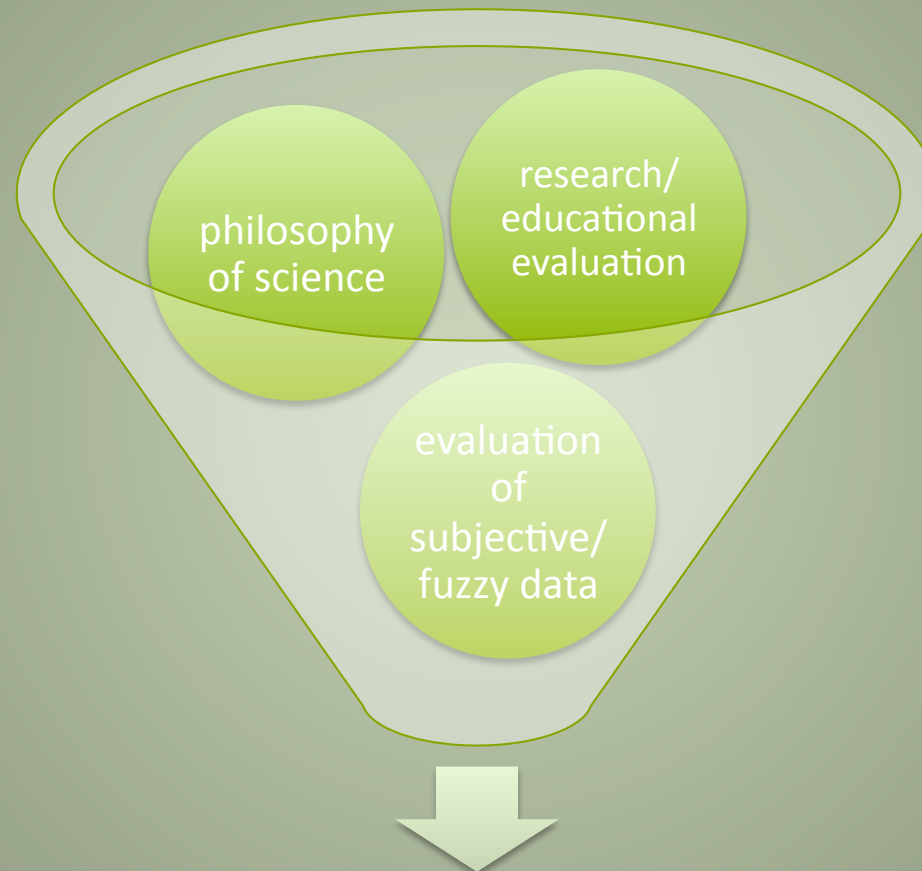


Jordanous (2012) – SPECS



Asking people how creative the system is

Which evaluation method is best for evaluating **how** creative our systems are



cross-disciplinary review of good practice in evaluation

Creativity meta-evaluation standards





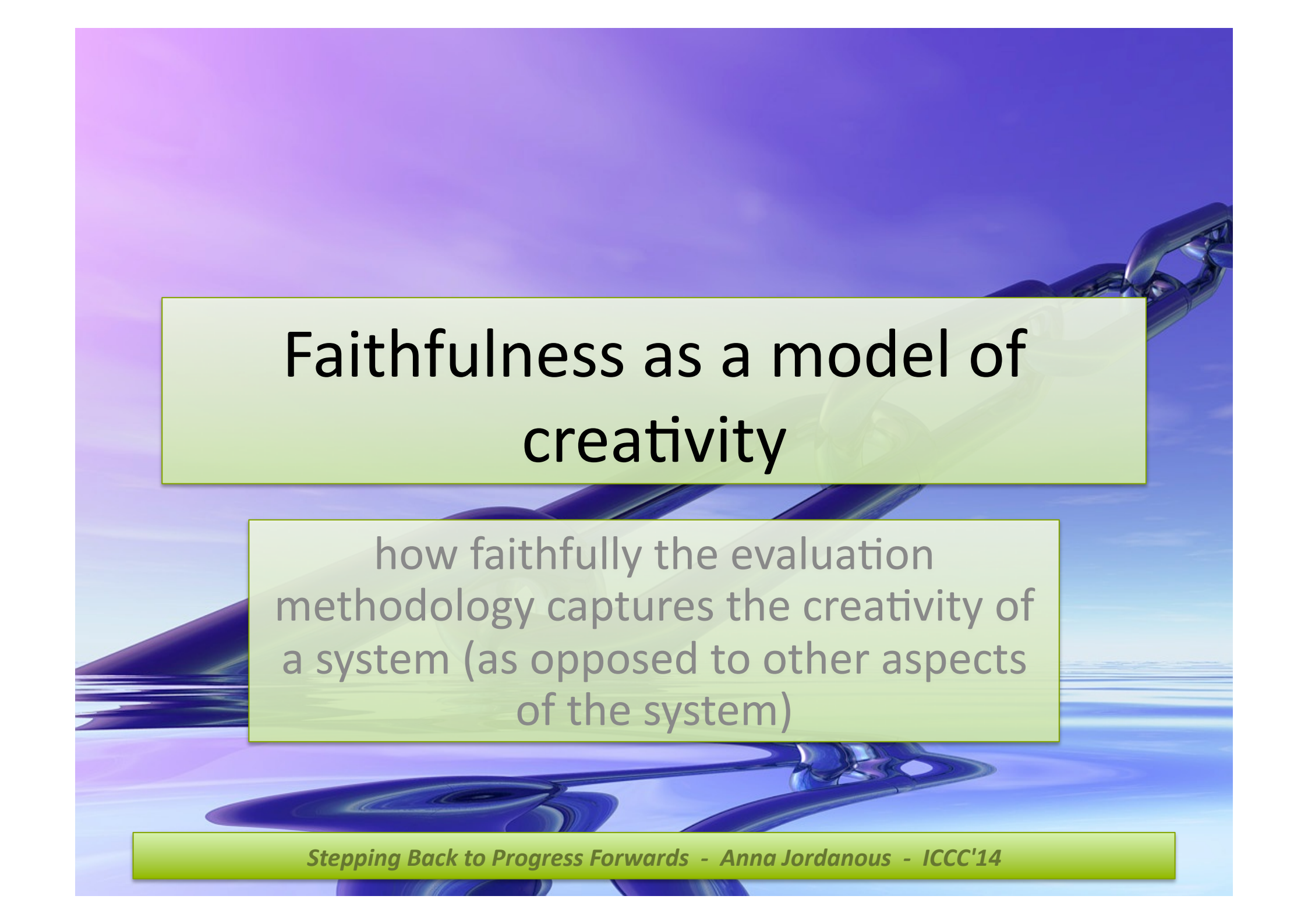
Correctness

how accurately and comprehensively
the evaluation findings reflect the
system's creativity



Usefulness

how informative the evaluative findings are for understanding and potentially improving the creativity of the system



Faithfulness as a model of creativity

how faithfully the evaluation
methodology captures the creativity of
a system (as opposed to other aspects
of the system)



Usability of the methodology

the ease with which the evaluation methodology can be applied in practice, for evaluating the creativity of systems



Generality

how generally applicable this methodology is across various types of creative systems

Testing this out: case study on evaluating creativity



Method

- Five evaluation methods used to evaluate creative musical improvisation systems
 - Ritchie's empirical criteria
 - SPECS (+cc)
 - Colton's creative tripod model
 - FACE model
 - Opinion survey
- Results considered by external evaluators along each meta-evaluation criterion and overall
 - evaluators = authors of the evaluated music impro systems

Results

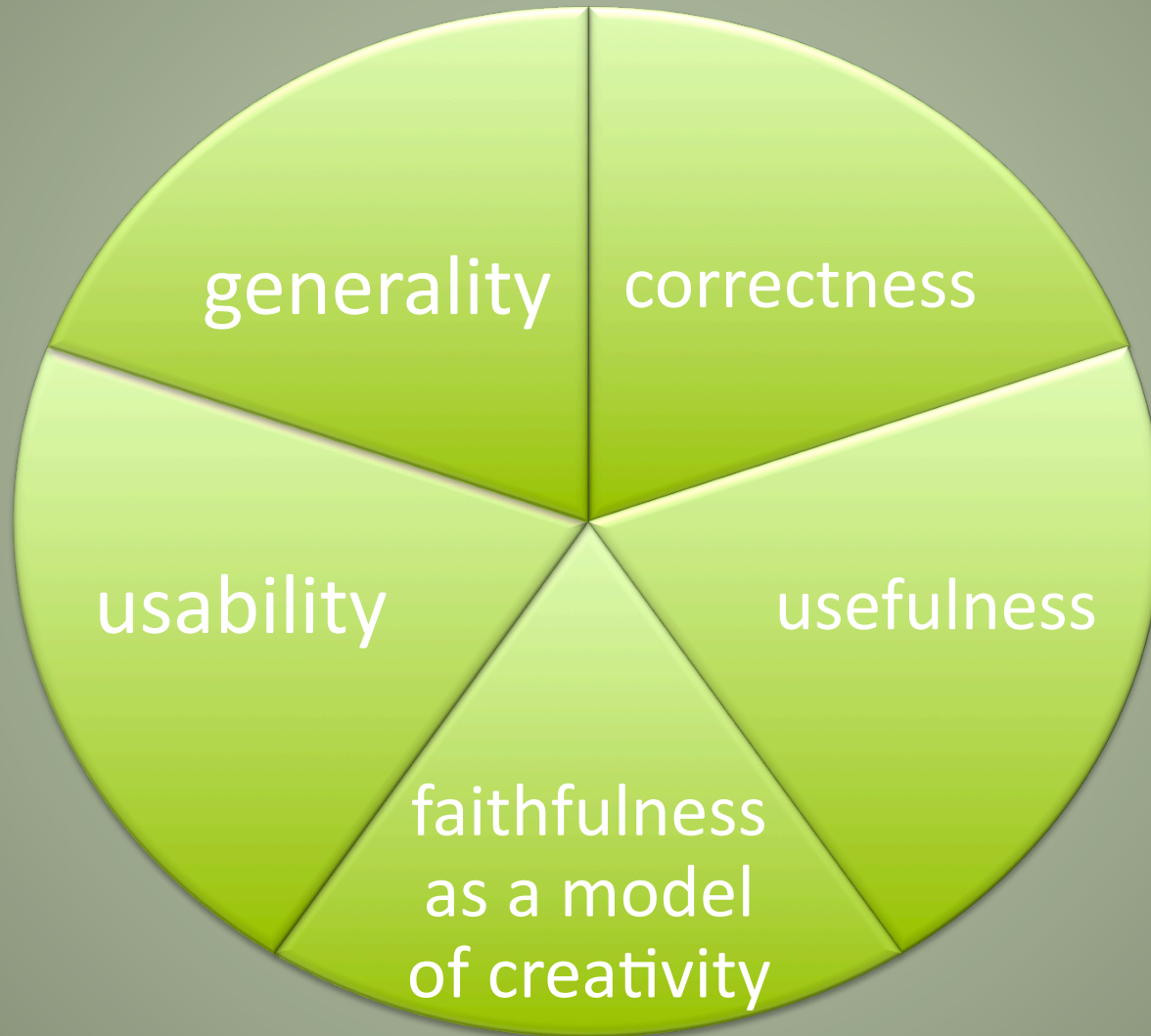
- + SPECS+cc performed well on most of the five meta-evaluation criteria
 - volume of data produced → **usability** issues
- + Colton's creative tripod was **the easiest to use**
 - some concerns about the **generality** of the tripod across creative domains & its **faithfulness** modelling creativity
- + Ritchie's criteria considered **accurate**
 - **usability** issues with abstract nature of the criteria and accompanying function definitions.
- + The FACE model considered quite **user friendly**
 - limited in how it dealt with aspects of creativity that were important but not in the FACE model.
- Each evaluation methodology was an improvement over the opinion survey

Overall: SPECS+cc and Ritchie's empirical criteria did best

Warning: One size does not fit all



Summary



(warning: one size does not fit all)

