

## How do we know that a system works?

- A. The serendipity Airbus? cf a "what if?" design for an Airbus  
explanation as knowing the relation between parameters and performance  
cf Bob Spence's artefacts to animate relationship between parameters and performance  
what is the status of cause and effect?
- B. By what criterion does a system work?  
Whose perspective? In general need to take account of many perspectives.  
Can't expect precise correspondence between behaviours of complex systems  
Need for method of decomposition and a compositional approach  
Empirical primitive components
- C. What is meant by an explanation from theory?  
Possible candidate: reductionism - a **Theory of Everything**  
Stewart and Cohen critique: **The Collapse of Chaos** "chains of logic too great"  
cf Langton's ant, how to infer emergent behaviour  
How does a mathematical theory accommodate experiment? how does it accommodate ignorance?  
How can mathematical descriptions deal with non-functional requirements?  
cf Russell's exploration of the agent-less world as the most appropriate world view,  
and Hoare's view that concurrent systems are best understood in agent-less terms
- D. Attribution to agents as an alternative to reductionism:  
don't venture to reduce everything to fundamental laws: cf how compass behaves but not "why"  
also allows for free-will: I act this way because I choose to do so - cf Pylyshyn
- E. The knowledge requirements for **system design framework**  
"what if" not just "what is": hence what is believed vs. what is observed  
the empirical-theoretical pyramid of explanation  
sections of **this pyramid as partial designs**  
personal -> **public, particular** -> general, experimental -> conviction

### Sources

Agent-oriented Modelling and the Explanation of Behaviour  
Gooding, Cohen and Stewart, von Neumann