Quality, Measurement (and Assurance Cases)

Tom Maibaum
McMaster Centre for Software Certification
tom@maibaum.org

McSCert

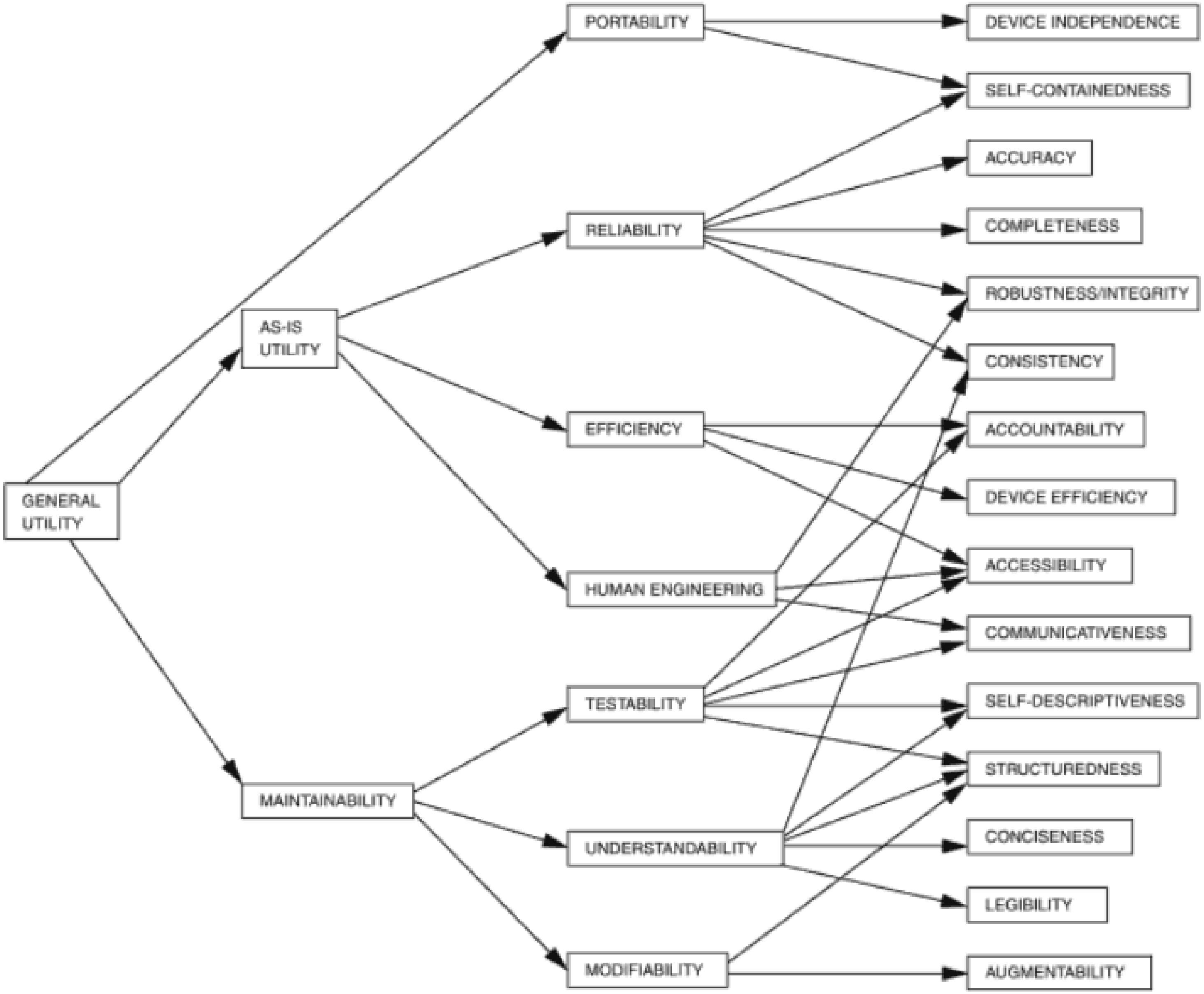
Quality Standards and Safety

- What is the relationship between quality models, quality standards (like SQuaRE) and assurance cases?
- Answer:
  - A quality model "is" a measurement framework, like IS for physical sciences
  - Existing quality models are very poor substitutes for proper measurement frameworks (so standards are commensurately bad)
  - Intuition is not a proper basis for engineering
  - Even a "proper" quality model (or corresponding standard) is *absolutely not* an assurance case for anything, including quality!

SI units

Fundamental quantity	Symbol for the quantity	Base unit	Symbol for the unit
mass	m	kilogram	kg
length	l	metre	m
time	t	second	s
temperature	T	kelvin	K
electric current	I	ampere	A

Derived quantity	Unit	Symbol for unit	Derivation
acceleration	metre per second squared	$m\,s^{-2}$	
area	metre squared	m^2	
density	kilogram per metre cubed	$kg\,m^{-3}$	
electric charge	coulomb	C	$1\,C = 1\,A\,s$
energy	joule	J	$1\,J = 1\,N\,m$
force	newton	N	$1\,N = 1\,kg\,m\,s^{-2}$
momentum	kilogram metre per second	$kg\,m\,s^{-1}$	
potential difference	volt	V	$1\,V = 1\,J\,C^{-1}$
power	watt	W	$1\,W = 1\,J\,s^{-1}$
pressure	pascal	Pa	$1\,Pa = 1\,N\,m^{-2} = 1\,kg\,m^{-1}\,s^{-2}$
velocity	metre per second	$m\,s^{-1}$	
volume	metre cubed	m^3	



Metrics

Quality model: Boehm (Selby, 2007)

Models, Measurement and Quality

McSCert

- Life gives rise to problems, calling for *decisions*.
- In everyday life:

  experience, common sense, intuition ⇒ solution
- Inadequate (though still essential) for solving managerial, technological, scientific problems of modern society.
  - experience and practices of today may be obsolete tomorrow
  - problem is so complex that intuition alone is not enough
  - solutions may not be derivable from the problem solver's experience alone
  - margin between good and bad decisions may be too narrow for informal treatment
  - value judgements may need *explicit justification*

Measurement

McSCert

- Scientific results and engineering achievements can only be judged on the basis of evidence.
- Convincing evidence can only be provided by *measurement*:

  "Measurement is the key to all disciplines of science and technology, and the maturity of the discipline is marked by the extent to which it is supported by a sound and comprehensive system of *measures*, *measurement standards*, *measurement tools* and *measuring procedures*."
- Measurement is the basis of expressing values and forming judgements.
- Measurement is the instrument by which to assure the *quality* of products and services (and the fairness of business and trade).

Measures

"Nothing exists in itself. There is no quality in this world that is not what it is merely by contrast." *Herman Melville*, *Moby Dick*

"What is not measurable, make measurable." *Lord Kelvin*

Measurement assists the decision maker in:

- problem formulation
- observation
- informed comparison
- reasoned discrimination

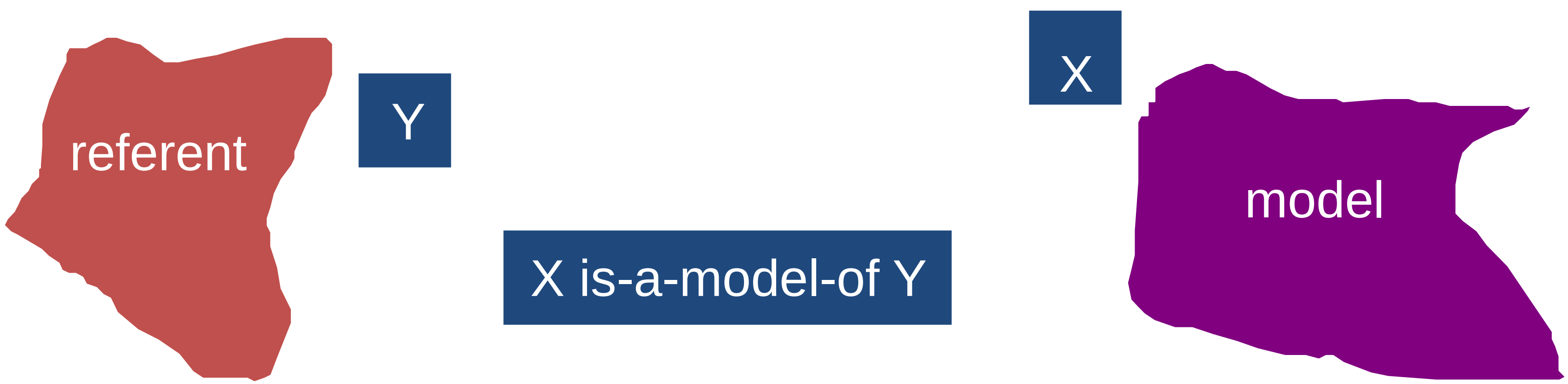
McSCert

What is Measurement?

- purpose of measurement is to provide a valid, trustworthy, traceable representation of some chosen *entity* whose selected *attributes* are of interest
- measures *may be* (obviously!) quantitative (but data gathering and casual assignment of numbers to things do not constitute measurement)
- some very important measures may be non-numerical
  - blood types
  - correctness of SW
  - alarm conditions
  - quality level

Referents, Attributes and Properties

- The basis of science and engineering is the construction of *models*.
- *Modelling* is essential for measurement: the purpose of modelling is to delimit the aspects (or parts) of the *referent* considered to be of interest (or necessary for solving the problem at hand).
- For our purposes, a model is a kind of *representation*:



McSCert

Referents, Attributes and Properties

- REFERENT
  entity in the real world; multitude of (empirical)
  - attributes
  - properties

- MODEL
  abstraction serving some engineering purpose; defined in terms of some interesting
  - attributes
  - properties

  which may be either *objectively observable* **or** *subjective* **!!**

Referents, Attributes and Properties

McSCert

A *measure* is a particular value of a *property variable* (used to model an *attribute*).

To characterise a referent, a measure must be assigned to each property which corresponds to its (<u>objective</u> or <u>subjective</u>) attributes.

Measures may be:

- QUANTITATIVE: use *symbol systems* which are complex
- QUALITATIVE: use *symbol systems* which are small discrete domains

Measures

Measurement is a 2 step process:

- a class of referents is modelled to represent a common attribute by a property variable (so as to compare referents via values of this attribute)
- a measure is assigned to the property variable of each item, so that the comparison of measures reflects comparisons about the referents

Usually, a referent is modelled by a set of attributes with well defined relationships between them (e.g., as a structure over a first order language)**.**

Measures

MEASUREMENT is the process (aka procedure) of making empirical observations about referent entities and representing their properties in a symbol system, so as to describe them.

A *MEASURE* is a product of the measurement process: it is a symbol of the symbol system designating the value of a property of the referent.

A *direct* measure is one which may be determined by direct observation (e.g., number of lines of code).

An *indirect* measure is one which is derived in some formal manner from other (possibly direct) measures (e.g., density in terms of weight and volume, fault density in terms of lines of code and number of faults observed).

Measures

UTILITY MEASURES

- The ultimate aim of measurement is to assist choice and to support decision making.
- Choice is always subjective.
- Rational, *informed* choice in science, technology or business relies on fact (and should not be random or capricious).

PROFESSIONAL ACCOUNTABILITY requires that the factual basis of the choice should be defined, the value system be explicit and the decision repeatable.

Utility Measures

- One must construct an explicit model of the subjective attribute on which the judgement is made.
- The *utility property* is given as a function of the directly or indirectly measurable objective properties of the referent.
- The *arguments* of the function are measures of objective attributes, reflecting ‘facts’.
- The *form* of the function is subjectively determined by the problem solver, reflecting judgement.
- The *value* of the function, the *utility measure*, is subjective, but is explicitly defined, its further use objective.

Characterising Measures

McSCert

Some laws/axioms of measurement theory:

- REPRESENTATION CONDITION

  A set of measures is a valid *representation* of a referent with respect to a given attribute if the mapping from the empirical domain of attributes to the formal domain of measures is a homomorphism. The relation (of representation) is irreflexive, asymmetric and transitive.

Characterising Measures

- UNIQUENESS AND SCALING:

  The scale (i.e., symbol system) chosen must either be unique or the truth value of a statement must remain invariant under all admissible transformations. (So the measurement scale adopted is in general not unique for the purpose at hand.)

  A measurement statement is said to be *meaningful* if its truth value is invariant under all admissible transformations.

  So, is the following meaningful?

  The temperature in Washington today is twice the temperature in Toronto.

Classical Scales

- Nominal scale: (classification)
  - e.g., blood groups, programming language, colour
- Ordinal scale: (ordering)
  - e.g., {Excellent, Very Good, Good, Satisfactory}
- Interval Scale: (quantifying differences)
  - e.g., Date. Temperature (except Kelvin)
- Ratio scale: (ratios and zero are meaningful)
  - e.g., Length
- Absolute scale: (counting)

Scale Types Summary

Scale Types	Characteristics
Nominal	Entities are classified. No arithmetic meaningful.
Ordinal	Entities are classified and ordered. Cannot use + or -.
Interval	Entities classified, ordered, and differences between them understood (‘units’). No zero, but can use ordinary arithmetic on intervals.
Ratio	Zeros, units, ratios between entities. All arithmetic.
Absolute	Counting; only one possible measure. All arithmetic.

Measuring quality attributes

McSCert

- What are the appropriate symbolic scales for measuring quality attributes?
  - Multiple scales for different attributes
- Is overall quality a “single” value, or a tuple of values?
  - Depends on how fine a judgement needs to be made
- What is the appropriate concept of threshold?

Measuring quality(, safety, assurance case confidence…)

- Define a model of quality suitable for your purposes, i.e.,
  - Define the attributes in terms of which quality is to be characterised
  - Organise them into a definitional hierarchy of base units and derived units
  - Assign appropriate measurement scales to these base and derived units
  - Define appropriate measurement procedures for base attributes, utility functions for (some) derived attributes
  - Validate model empirically, checking satisfaction of measurement laws
  - Train engineers in use of measurement framework
- Now use the measurement result as evidence in your assurance case

Quality of assurance cases: Ripeness for evaluation

- Readiness for submission: is there a sufficient weight of evidence to submit the case for evaluation?
- (Is the evidence in a criminal investigation of sufficient weight to warrant a trial?
  - NOT THE SAME JUDGEMENT AS "IS THE ARGUMENT CONVINCING DURING A TRIAL"!)
- We are investigating the measurement of confidence in assurance cases.
- Having sufficient weight of evidence (ripeness) is a kind of quality measure for an assurance case.

McSCert

Quality of assurance cases: Ripeness for evaluation

Confidence in an assurance case is the quality or state of being certain that the assurance case is appropriately and effectively structured, and correct.

- A working definition that provides the hooks for defining a candidate measurement framework for confidence.
- Ripeness is a kind of quality measure associated with the assurance case.
- Measurement scales for confidence are Baconian, not statistical or Bayesian.

Scales for confidence and ripeness

- Two notions of weight:

"the degree to which a rational decision-maker is convinced of the truth of a proposition [actually plausibility of its proof!!!] as compared to some competing hypothesis (which could be simply that the proposition is false)" [Nance] CONFIDENCE

"a balance, not between the favourable and the unfavourable evidence, but between the *absolute* amounts of relevant knowledge and relevant ignorance. As the relevant evidence at our disposal increases, the magnitude of the probability of the argument may either decrease or increase, according as the new knowledge strengthens the unfavourable or the favourable evidence; but *something* seems to have increased in either case, we have a more substantial basis upon which to rest our conclusion." [Keynes]

Scales for confidence and ripeness

- We need some means for measuring, ranking and comparing weight. From Cohen:
  - "… the principle of equipollence is indefensible, [therefore] there is no natural unit of weight and the prospects of any non-arbitrary system for measuring weight are very poor"
  - Cohen suggests there is a way to rank and compare it, at least for arguments about a given subject-matter (domain).
  - To accomplish this, we need "… an ordering for a certain set of families of evidential predicates and [to concern] ourselves only with arguments from premisses [sic] that contain just predicates belonging to the first family, or just those predicates plus predicates belonging to the second family, or just predicates from each of the first three families, and so on cumulatively"
  - "… it will be important to give priority in the ordering of predicate-families to those families that contain at least one predicate which is highly relevant in relation to the accepted prior probability of at least one conclusion in the given field"

A lot of hard work ahead!

- Given a domain which is "narrow enough":
  - Identify what "quality" is to be measured
  - Enumerate all potential sources of evidence and and their comparative relevance
  - Define a measurement framework for it, including all relevant base measures and derived measures (some using utility functions)
  - Assign *a priori* weights to basic ones (based on best practice)
  - Define weights for derived measures (based on best practice)
  - Adjust in light of experience with framework, identification of new sources of evidence