Knowledge Elicitation Techniques in RE

**Background**
- Knowledge elicitation is concerned with discovering 'expert' knowledge
  - Grew out of Expert Systems work in the 80's
  - Originally focussed on deriving expert's "rules" for Rule-based Systems
  - More recently, focussed on "problem solving methods"

**But KE is hard**
- Separation of domain knowledge from performance knowledge
- Modeling problems
  - Britteness
  - Assumption of rationality
- Representational Problem
  - epistemological inadequacy
  - expressiveness vs. acquirability
- Expert Bias

**Example Techniques**
- Eliciting domain knowledge
  - Card Sorting
  - Laddering
  - Proximity Scaling Techniques
- Eliciting performance knowledge
  - Protocol Analysis
- Using Multiple Experts
  - Delphi Technique
  - Focus Groups
  - Repertory Grids
- Automated Techniques
  - Machine Learning
Why is KE so hard?

- Experts are not used to describing what they do.
  - Three stage model of learning:
    1) cognitive - verbal rehearsal of tasks;
    2) associative - reinforcement through repetition, verbal mediation disappears
    3) autonomous - compiled, no conscious awareness of performance.
  - Procedural and declarative are different mechanisms
    - Procedural knowledge becomes procedural with repeated application - experts lose awareness of what they know and cannot introspect reliably
    - Experts have little or no introspective access to higher order cognitive processes

- Representational Problems
  - Experts don’t have the language to describe their knowledge
    - No spoken language offers the necessary precision
    - Knowledge Engineer and Expert must work together to create a suitable language
  - Different knowledge representations are good for different things
    - Epistemological adequacy: does the formalism express expert’s knowledge well?

- Britteness
  - Knowledge is created, not extracted.
    - Knowledge models are abstractions of reality and hence are unavoidably selective
    - Britteness caused by the simplifying assumptions - instead of adding more knowledge, a better (more comprehensive) model is needed.

Expressiveness vs Acquireability
The Knowledge Level

- View knowledge modelling as:
  - Observe behaviour of an agent as black box
    - It acts as if it has some knowledge about its environment which it uses rationally
    - It takes actions to achieve ascribed goals
  - Construct two models:
    - Symbol Level - descriptions for mechanising behaviour
    - Knowledge Level - descriptions of the agent’s knowledge of the world

Two-step rationality:
- Agent applies its knowledge in two stages:
  - First creates a task specific model from the KL model based on features of the task.
- Hence, we actually need 3 models:
  - Domain model - a systematic way of talking about a domain, with a coherent ontology.
  - Task model - models goals, what it means to achieve a goal, and how goals are related.
  - Problem-solving method - a way of relating task and domain models to accomplish goals.

Knowledge Elicitation Techniques

- Protocol Analysis
  - based on vocalising behaviour
    - Think aloud vs. retrospective protocols
  - Advantages
    - Direct verbalisation of cognitive activities
    - Embedded in the work context
    - Good at revealing interaction problems with existing systems
  - Disadvantages
    - Essentially based on introspection, hence unreliable
    - No social dimension

- Proximity Scaling Techniques
  - Given some domain objects, derive a set of dimensions for classifying them:
    - step 1: pairwise proximity assessment among domain elements
    - step 2: automated analysis to build multi-dimensional space to classify the objects
  - Advantages
    - help to elicit mental models, where complex multivariate data is concerned
    - good for eliciting tacit knowledge
  - Disadvantages
    - Requires an agreed on set of objects
    - Only models classification knowledge (no performance knowledge)
## more KE techniques

### Card Sorting
- **Card Sorting**
  - For a given set of domain objects, written on cards:
    - Expert sorts the cards into groups...
    - Then says what the criterion was for sorting, and what the groups were.
  - **Advantages**
    - Simple, amenable to automation
    - Elicits classification knowledge
  - **Problems**
    - Suitable entities need to be identified with suitable semantic spread across domain.
    - No performance knowledge

### Laddering
- **Laddering**
  - Uses a set of probes (types of question) to acquire structure and content of stakeholders' knowledge.
    - Interview the expert.
    - Use questions to move up and down a conceptual hierarchy
  - **Advantages**
    - Deals with hierarchical knowledge, including poly-hierarchies (e.g., goal trees, “is-a” taxonomies).
    - Knowledge is represented in standardised format
    - Can elicit structural knowledge
    - Suitable for automation.
  - **Disadvantages**
    - Assumes hierarchically arranged knowledge.

### KA from Multiple Experts

#### Delphi technique
- **Delphi technique**
  - Used where contact between experts is difficult:
    - Each expert submits their judgement
    - All judgements are circulated anonymously to all experts
    - Each expert then submits a revised judgement
    - Iterate until judgements converge

#### Focus Groups
- **Focus Groups**
  - A technique derived from marketing:
    - Assemble experts together and discuss the problem
    - Discussion may be structured (e.g., debate) or unstructured

#### Repertory Grids (based on Kelly's Personal Construct Theory)
- **Repertory Grids**
  - Used to detect terminological differences
    - Get the experts to agree a set of entities
    - Each expert provides attributes and values
    - For each attribute in expert A’s grid, find the closest match in expert B’s grid.
      (i.e. are there attributes which have the same discriminatory function?)
    - Experts then rate the entities using each other’s attributes
Abstractionism vs. Contextualism

Abstractionism
- Builds models abstracted from a domain; the model is used to answer questions
  1. Decide on the ontology of the phenomena we wish to describe
  2. Use this ontology to represent the domain of discourse
- Assumes knowledge and understanding are independent from context
- Used by natural scientists and engineers.
  - Although many scientists don’t realize that step 1 involves choice
  - Logical positivism vs. theory-driven observation

Contextualism
- Emphasizes the details and idiosyncrasies of the domain
  1. Collect naturalistic data from the domain of study (Rich descriptions)
  2. Use the data to support explanations (but don’t build abstract models)
- Assumes it is impossible to build models that have meaning when removed from their context
- Used by many social scientists
  - But generally limits them to the descriptive rather than predictive/prescriptive

Participant Observation

Approach
- Longitudinal studies:
  - Observer spends time with the subjects, joining in, long enough to become a member of the group

Advantages
- Contextualized;
- Reveals details that other methods cannot

Disadvantages
- Extremely time consuming!
- Resulting ‘rich picture’ is hard to analyze
- Cannot say much about the results of proposed changes

Watch for
- Going native!
Ethnomethodology

**Basis**
- Social world is ordered
  - The social order may not be obvious, nor describable from common sense
- The social order cannot be assumed to have an a priori structure
  - Social order is accomplished on a moment-to-moment basis through participants’ collective actions (rather than through any pre-existing structures)
  - i.e. social order only observable when an observer immerses herself in it.
- Observation should be done in a natural setting
- Need to consider how meanings develop and evolve within context

**“Use the members’ own Categories”**
- Most conventional approaches assume preexisting categories
  - This may mislead the observer (e.g. appropriation)
- Ethnography attempts to use the subjects’ own categories
  - What categories (concepts) do they use themselves to order the social world?
- What methods do people use to make sense of the world around them?
  - Use the same methods members use during observation, for example, by developing a legitimate role within the community under observation.

**Measurement**
- No scientific objectivity, so use the subjects’ own measurement theory

---

Ethnomethodological approach

**Ethnomethodology is a subarea of Anthropology**
- Looks for behaviours that may be different in a specific culture but which have the same underlying purpose or meaning.
  - E.g. how do people go about gaining status in different cultures:
    - Frenchmen brag about sexual conquests to gain status;
    - Americans brag about money to gain status.
  - Each of these topics is taboo in the other culture

**Uses a very tightly controlled set of methods:**
- Conversational analysis
- Measurement of body system functions - e.g. heartbeat
- Studies of Non-verbal behaviour (e.g. gestures, body language)
- Detailed video analysis
  - These techniques are useful in capturing information about a social setting.

**Other observation techniques can be applied:**
- Time-motion study
  - who is where, when?
- Communication audit
  - who talks to whom about what?
- Use of tools - status symbols plus sharing rules
Postscript: Postmodernism

Modernism
- Rationality is the highest form of mental functioning
- Modern science produces universal truths
  - independent from the context and status of the scientist who produced them
- Rationality will always lead to progress and perfection
  - All human institutions can be scientifically analyzed and improved
- Reason is the ultimate judge of what is right (true, legal, ethical, ...)
- Language must be rational
  - it only exists to represent the real world; there must be a firm, objective connection between the "signifier" and the "signified"
  - the meaning cannot depend on the audience

Postmodernism
- Questioning the grand narrative
  - A grand narrative is a story that a culture/society tells itself about its practices and beliefs
  - E.g. in the US: "democracy is the most enlightened/rational form of government"
  - E.g. in science: "scientific truths are universal and eternal"
  - Postmodernism identifies and critiques such narratives
- Instead, look for mini-narratives
  - Stories that explain small practices, local events, situated, contingent behaviour
  - and don't make any claims about universality, truth, or stability
- E.g. Literary Deconstruction
  - Examine what a text does not say, what it represses
  - Reveal internal arbitrary hierarchies and dichotomies
- E.g. Semiotics
  - The study of the relationship between signs and the things they signify

An RE Methods Classification (after Lyotard)

Modern
- assume a single objective reality
- Hard: an organization is a rational system
- Soft: a system can serve multiple objectives

Post-Modern
- allow for multiple views within society
- Pluralistic
  - Societies are based on "local language games" and cannot be unified or neatly divided into parts
- Divisive
  - a system can serve multiple objectives
- Dual
  - based on the marxist conflict
  - RE must take sides
  - Critical
    - seek alternatives to existing social conditions
  - Democratic
    - seek to involve all viewpoints in a democratic style
  - Network
    - evolutionary approaches

Source: Adapted from http://www.colorado.edu/English/ENGL2012Klages/pomo.html