CogSys
Cognitive Systems

COOPERATION

CLASSiC

www.classic-project.org

Automatic Learning and Evaluation of User-Centered Objective Functions for Dialogue System Optimisation

Verena Rieser and Oliver Lemon

School of Informatics

University of Edinburgh

TALK www.talk-project.org
(EC FP 6)

Outline

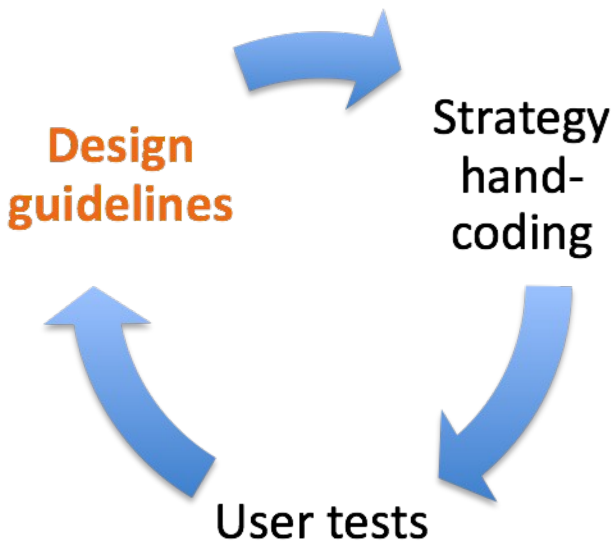
- **Area:** Dialogue strategy design and optimization via data-driven statistical learning
- **Problem:** Modelling “true” User Satisfaction

- **Techniques**: Reinforcement Learning
  - PARADISE regression models of user satisfaction (Walker et al. 1997, 2000)

- **Meta-evaluation**: comparing learned User Satisfaction models across 3 corpora

Dialogue strategy design methods

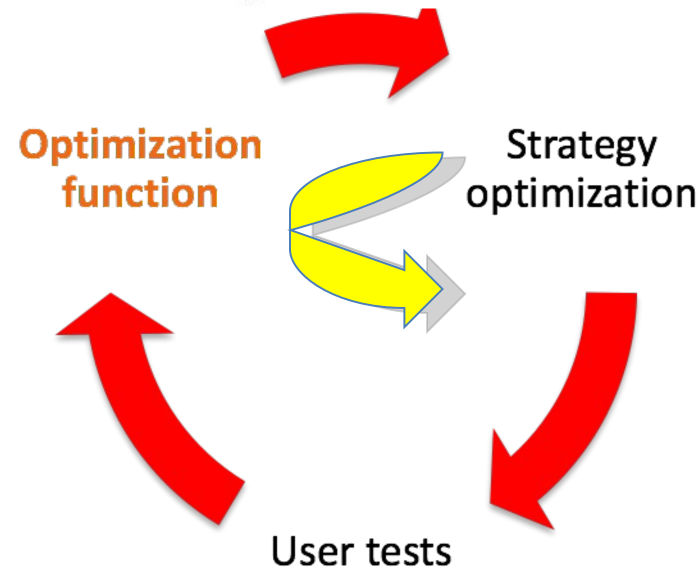
Conventional software life cycle



Design by `Best practices' (Paek 2007)

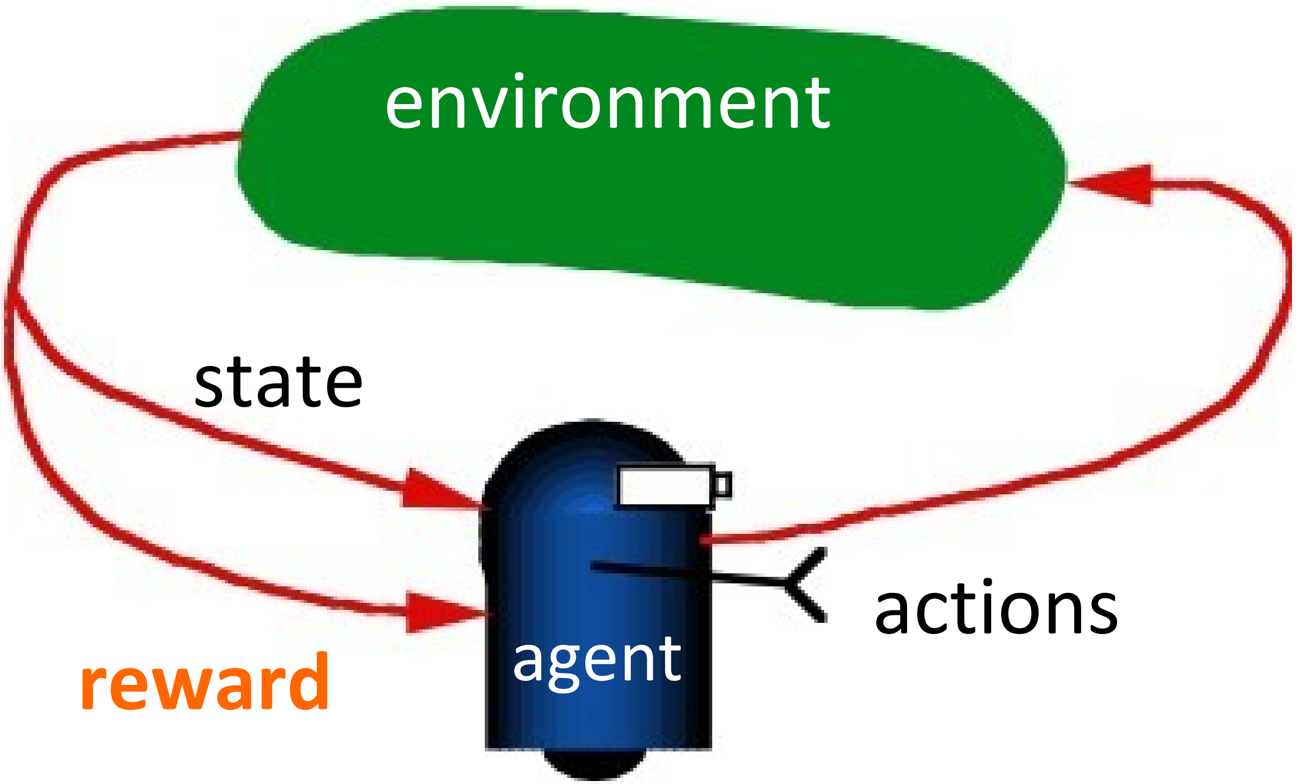
Automatic strategy optimisation

$$Q^{\pi}(s,a) = \sum_{s'} \mathcal{T}^{a}_{ss'} [\mathcal{R}^{a}_{ss'} + \gamma V^{\pi}(s')]$$



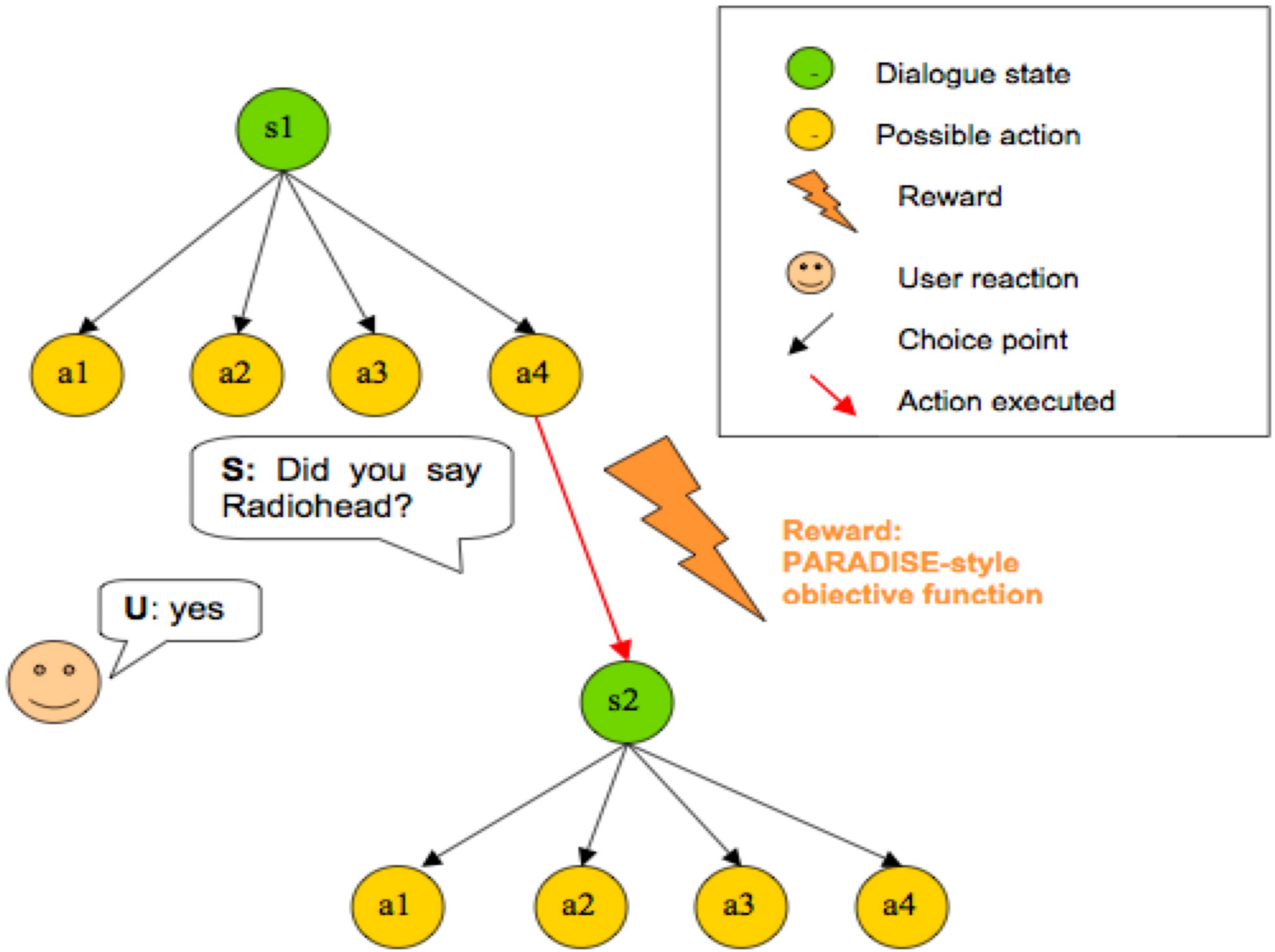
Automatic design by optimization function (= "reward")

Reinforcement Learning



$$Q^{\pi}(s,a) = \sum_{s'} \mathcal{T}^{a}_{ss'}[\mathcal{R}^{a}_{ss'} + \gamma V^{\pi}(s')]$$

Automatic Strategy Optimization using Reinforcement Learning



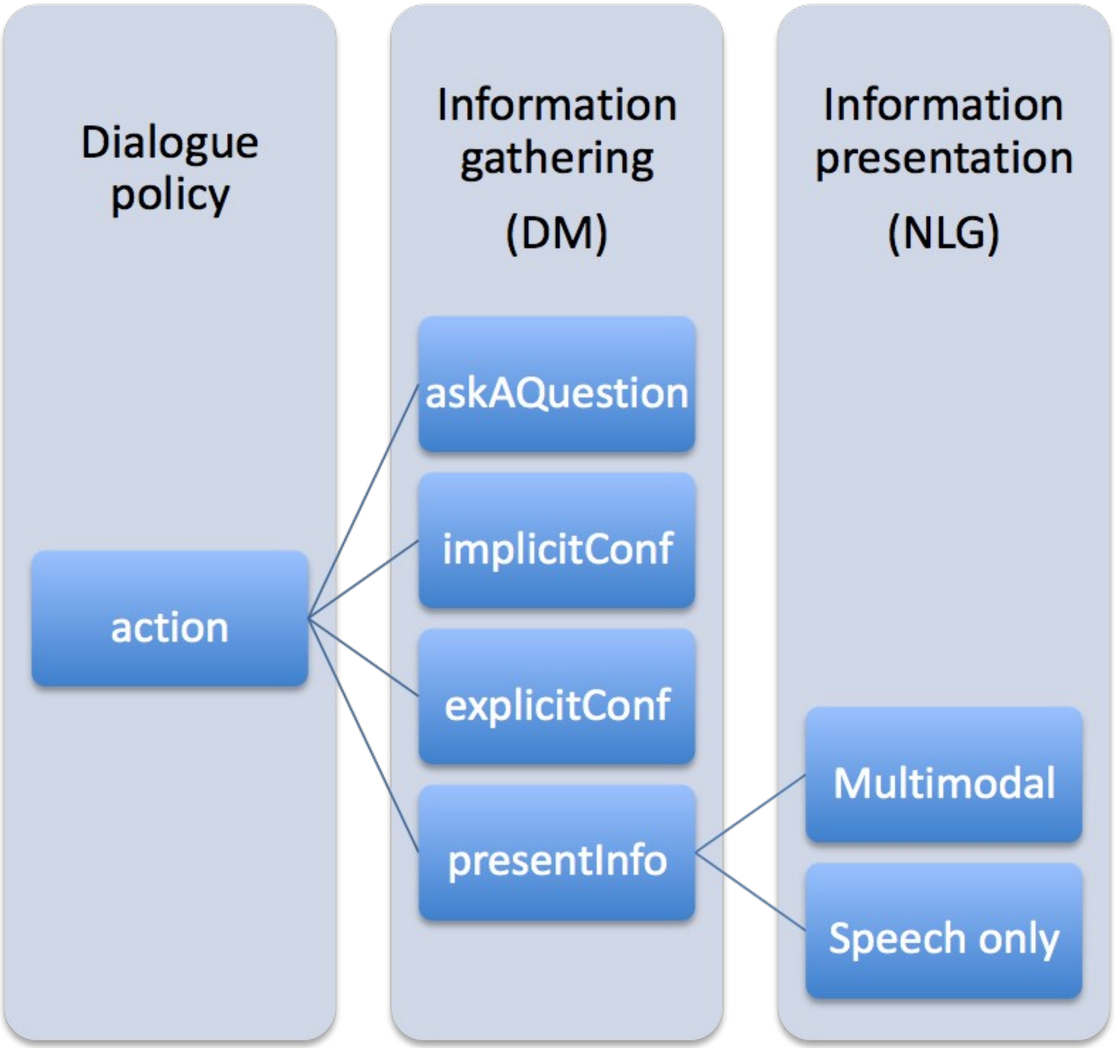
Problem: How to guarantee real User Satisfaction?

"Who picked 'I Can't Get No Satisfaction' to be our on-hold music?"

Research questions

- “Quality assurance” for reward/ objective functions: aim is to optimize for real user preferences
- Can we do better than:
  “Reward= Task Completion – Dialogue Length” ?
- “Bootstrapping”: is a reward function derived from a small Wizard-of-Oz data collection a valid estimate of real user preferences?

The decision/learning problem



Method

- Build simulated learning environment from data collected in a WOZ user study [Rieser & Lemon, ACL'06]

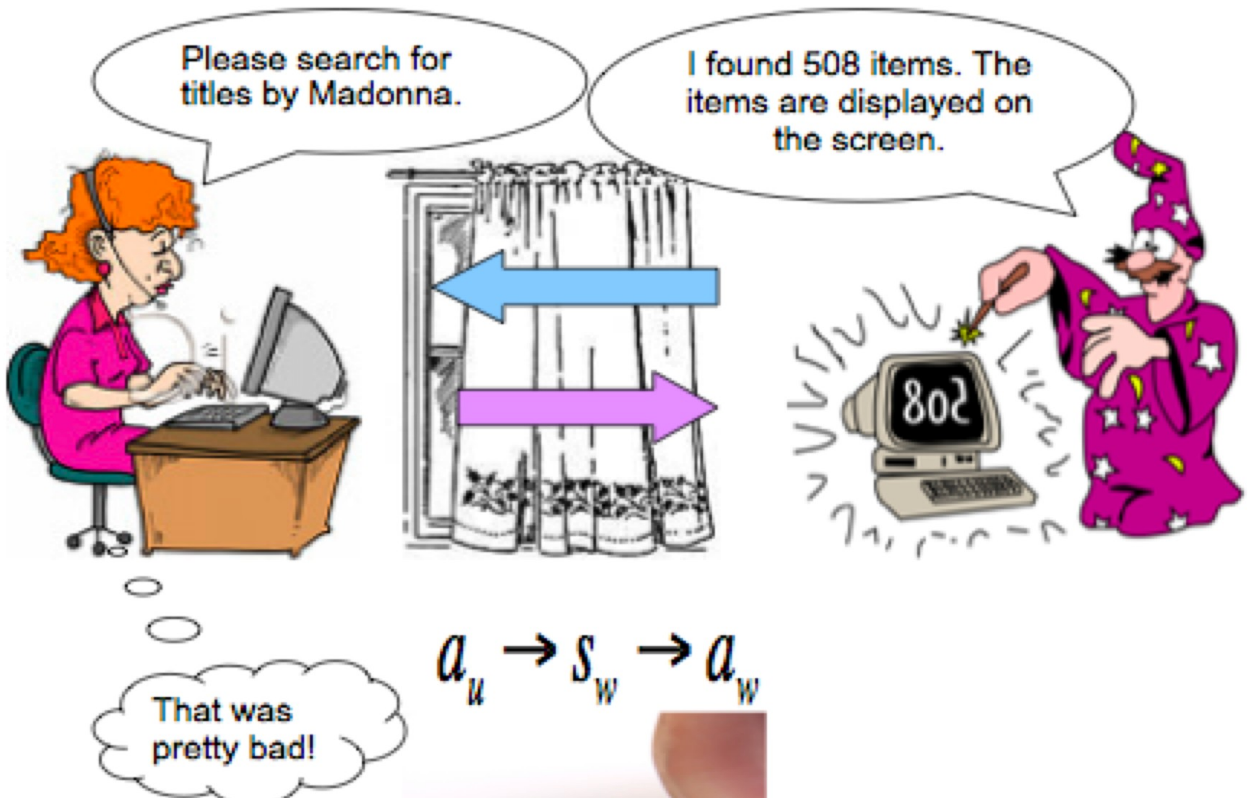
- Train and test RL policy by simulated interaction [Rieser & Lemon: Interspeech'07, JNLE'08]
  - Compare against supervised learning baseline policy (non-optimised policy)

- Test the 2 policies with real users (17 subjects)

  [Rieser & Lemon, ACL'08]

- Meta-evaluate: compare results from these 3 corpora [Rieser & Lemon, LREC'08] = this paper!

A Wizard-of-Oz experiment



$$a_u \rightarrow s_w \rightarrow a_w$$

- 5 point Likert scale

PARADISE evaluation framework

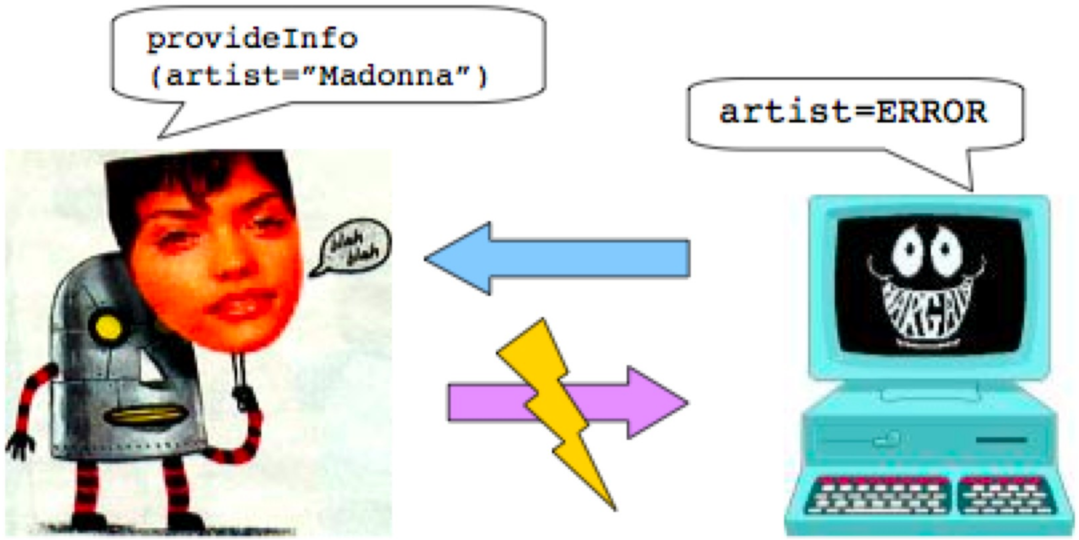
[Walker et al. 1997]

$$\underbrace{US}_{subjective} = \underbrace{\alpha \times \mathrm{N}(\kappa) - \sum_{i=1}^{n} w_i \times \mathrm{N}(C_i)}_{objective}$$

Automatic estimate of subjective **User Satisfaction** (US) from objective dialogue performance measures, using multivariate linear regression, where:

- κ : task success
- C_i : dialogue objective measures (e.g. dialogue length, Word Error Rate, number of confirmations........)
- w_i : weights assigned by regression

Use PARADISE reward model to train via Reinforcement Learning



$$a_u \rightarrow \tilde{s}_u \rightarrow \tilde{a}_s \rightarrow a_s$$

$$\underbrace{US}_{subjective} = \underbrace{\alpha \times \mathrm{N}(\kappa) - \sum_{i=1}^{n} w_i \times \mathrm{N}(C_i)}_{objective}$$

Problems/questions:

- Generality of PARADISE models across different systems and user groups (Walker et al. 2000, Paek 2007)?
- Performance of dialogue strategies optimized using PARADISE reward models?

Evaluation of the PARADISE model

- “Model stability”
  a. Build PARADISE model from WOZ data
  b. Build PARADISE model from real user test data
  c. Compare obtained models
- “Model performance” - prediction accuracy
  – Predict unseen events from the same data set
  – Predict unseen events from other data sets
- Performance of dialogue strategies learned with the PARADISE models

Model stability

TaskEase_WOZ = 1.58 + .12 taskCompl
+.09 mmScore -.2 dialogueLength

TaskEase_SL = 3.5 + .54 mmScore -.34 dialogueLength

TaskEase_RL = 3.8 + .49 mmScore -.36 dialogueLength

- Regression models show the same trends

Model performance: prediction accuracy

- Predicting unseen events in original system
- Predicting unseen events of new systems

- 10-fold cross validation on same data set
- Cross-system evaluations

- **Result:** Prediction accuracy is stable across the models and the systems (~16% error)
- The models generalize well

Performance in dialogue optimization

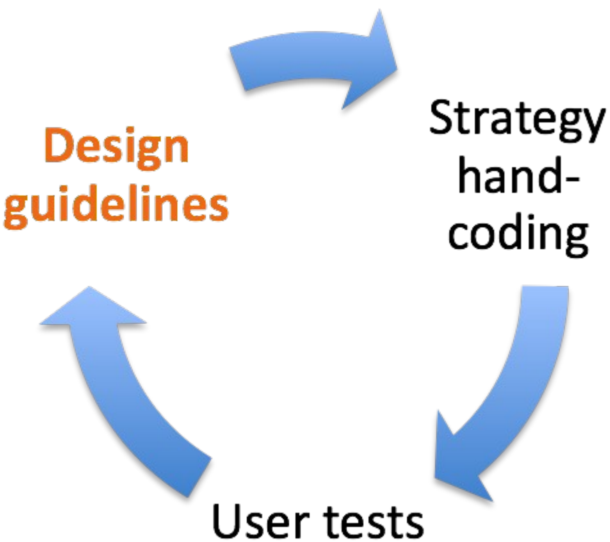
- 17 subjects, 204 dialogues (half SL, half RL)
- The RL policy significantly outperforms the non-optimised (i.e. SL) policy
  - 18 times more reward ($p<0.005$)
- Users rate the RL policy on average 10% higher ($p<.001$)
- See our ACL 2008 paper

Results

- **Overall:** method for design and optimization of dialogue systems ("bootstrapping")
- **This paper:** method for meta-evaluation of objective/ reward functions
- Despite learning from small amounts of initial WOZ data, a PARADISE-style objective function is a **stable, reliable, and useful** model of real User Satisfaction
- Moving dialogue system design from "art to science" (Sparck-Jones 1996)

Dialogue strategy design methods

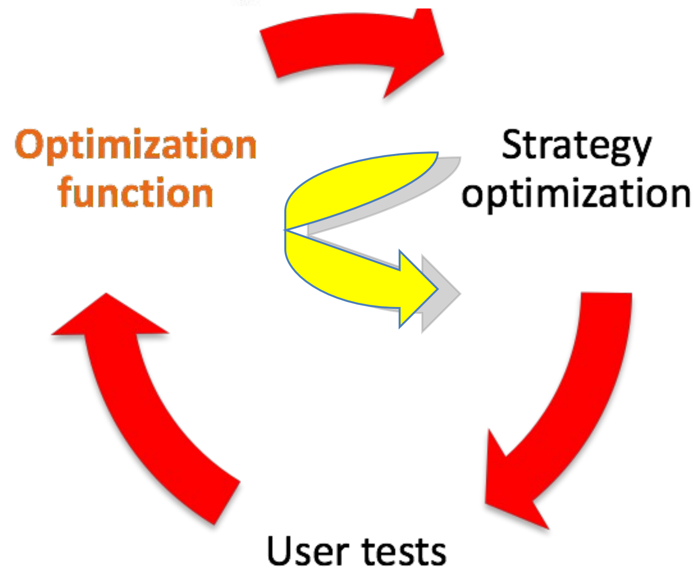
Conventional software life cycle



Design by `Best practices' (Paek 2007)

Automatic strategy optimisation

$$Q^{\pi}(s,a) = \sum_{s'} \mathcal{T}^{a}_{ss'} [\mathcal{R}^{a}_{ss'} + \gamma V^{\pi}(s')]$$



Bootstrap from WoZ data

Reward defined via PARADISE

Future work

- User preferences regarding **Natural Language Generation** decisions (see related papers at LONdial 2008)
- Incremental training with improved representations of user preferences
- More data! (e.g. From France Telecom/Orange Labs in the CLASSiC project)
- Further exploration of non-linear reward functions

Thanks for your time. Curious?

- See papers at: AISB MOG 2008, J. NLE 2008, LONdial 2008, ACL 2008
- See the TALK project www.talk-project.org (EC FP6)
- See the CLASSiC project (2008-11) “Computational Learning in Adaptive Systems for Spoken Conversation” (FP7 Cognitive Systems)
- www.classic-project.org

CogSys
Cognitive Systems

COOPERATION