Towards Human-Machine Symmetry (HMS)
(How to get AI into HCI?)

Aaron Sloman
http://www.cs.bham.ac.uk/~axs/misc/talks/#talk19

Currently there is huge asymmetry between computers and humans even when they interact using some structure that the computer understands, e.g. a form, a mouse-driven interface.

- In many interfaces there are things a human can perceive and understand that the computer cannot, even though it can manipulate them, e.g.
  - Instructions to the user
  - Questions to be answered
  - The answers typed in
  - Pictures and diagrams
- There are differences between what the machine and human can change in the display.
- Often the computer has too much control over what happens, and the human cannot do anything about this, e.g. asking to go back to an earlier stage in order to make a simple change, or wanting more space.
- Examples: editors, text processors, document formatters, graphical tools (xfig, tgif), email interfaces, compilers, software development environments, computer games, etc.

Human-Human interfaces

Compare two humans interacting using some physical display - paper, whiteboard, computer screen, email, etc.

- The medium is a neutral interface to which both have the same access

  Exceptions:

  - The two humans are in different locations
  - One has a disability e.g. can't see, can't write
  - One uses better tools than the other (e.g. for email)

- In general both can understand what is expressed in the medium

  (What are the exceptions?)

- In general if one cannot understand, the other can help, by giving an explanation or by rewording or re-drawing.

  (What are the exceptions?)

- Humans have affective states (desires, dislikes, preferences, attitudes, emotions, long term goals, immediate intentions) which can generate conflicts or a need/desire to cooperate, or influence attention.

- Humans can (sometimes) understand one another's affective states and take appropriate actions (being kind, tactful, vengeful, sarcastic, polite....)

What are the causes of the asymmetry?

What sorts of differences are there between humans and computers (programs)?

- They have different cognitive powers — what does that mean?
  - Their visual and other perceptual abilities are different.
  - They understand different languages – verbal and pictorial (e.g. varieties of syntactic forms understood – Examples??)
  - They use different ontologies,
  - They have different sorts of reasoning abilities,
  - They have different abilities to learn through doing and interacting,
  - They have different abililties to ask the other questions or explain things to the other
- They may have different physical abilities to access information:
  - The computer can't see the screen, usually (it has a very different 'view' of its contents). (What view?)
  - Humans can't see internal data-structures, usually
  - the human can't usually directly alter what's on the screen
  - The computer can't usually move the mouse or hit keys.
- They may have different goals and other affective states – or the computer may have no goals, desires, attitudes, etc., merely programs.

SOME DIFFERENCES MORE INTERESTING THAN OTHERS.
SOME ARE HARDER TO REMOVE THAN OTHERS.

How can we increase Human-Machine symmetry?

One option is to reduce functionality of the interface.

e.g. programmable editors, like Emacs and Ved.

- Limit the display to include only elements and structures that both the computer and the user can change
- Limit the display to include only elements and structures that both the computer and the user can read
- Limit the display content to something both human and machine can understand (limit semantic content to something both can understand) includes restriction of ontologies
- Limit goals and tasks to those either could in principle aim for and achieve

Alternatively, extend the computer's abilities, perhaps giving it new shallow behaviours.

- Increase cognitive and affective powers of computers (long term goals of AI)
- Meawhile attempt to disguise the machine's limitations - e.g. in so-called believable agents.

  I.e. produce superficial behaviour that gives the appearance of understanding, sympathising, trying to be helpful, sharing emotions, etc. (Compare computer entertainments).

Reducing functionality to increase symmetry

Some examples

- Ved and Emacs - considerable symmetry at the textual level extendable by user programs: unlike most editors
  - Program and user can write into the editor’s buffer
  - Program and user can read the editor’s buffer
  - Program and user can point in the editor’s buffer (moving the editor cursor).
  - However humans can still see larger scale phenomena that the computer cannot, and can interpret contents in ways the computer cannot.
    (Contrast text files and program files)

  Give demo

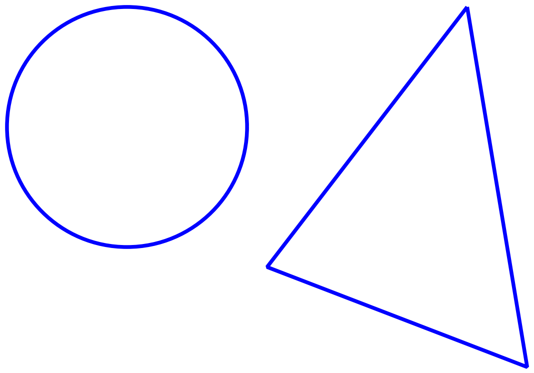
- RCLIB and SimAgent tools: Symmetry extended to graphical interface
  - Program and user can move objects around on the screen
  - Both may be able to create and delete objects
  - But there are still asymmetries in things they can manipulate

  Give demo

- But there is still a HUGE amount of work to be done.
- Does the asymmetry matter? Which kinds are most important?

We can think with diagrams

In this figure there are no points common to the triangle and the circle. Suppose the circle and triangle change their size and shape and move around in the surface.
They could come into contact.



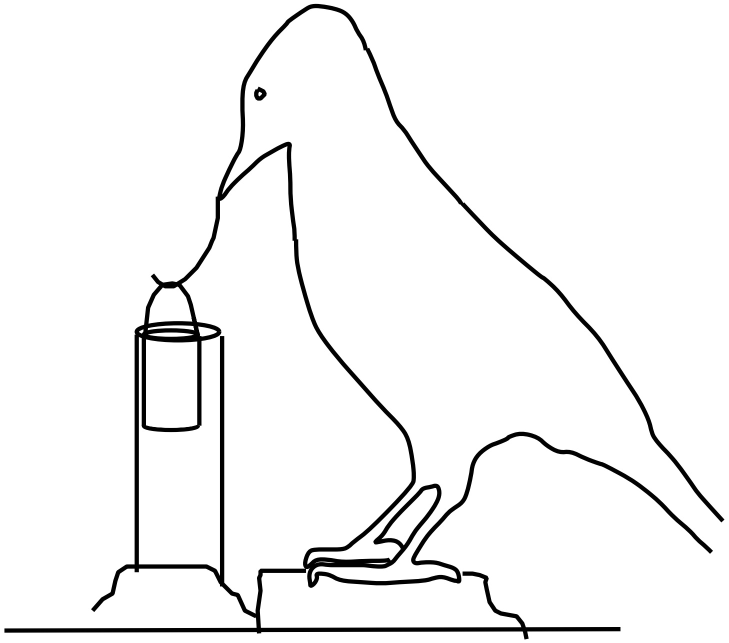
- **Clearly if a vertex touches the circle there could be one point common to both figures.**
- **If one vertex moves inside the circle and the remainder of the triangle is outside the circle how many points are common ot the circle and the triangle?**
- **What are all the possible numbers of common points?**

How do you try to answer the questions?

Can we give the computer the ability to perceive a structure and visualise possible changes in that structure?

Betty Crow: Cognitive Agent and Hook-maker

Two crows, Betty and Abel, learnt to use bent wire to fish a bucket of food out of the vertical tube (as in the picture). Then Abel flew off with the hook.



See the video here: http://news.bbc.co.uk/1/hi/sci/tech/2178920.stm
To find more, give google: betty crow hook

- Betty tried using a straight piece of wire for a while, and failed.
- She then pushed one end of the wire into the tape holding the tube and moved the other round with her beak, making a hook, which she used to lift the bucket.
- She did this 9 times out of 10. Reported in Nature and shown on BBC TV (August 2002).

COULD A ROBOT REPLICATE BETTY'S MENTAL PROCESSES?

What sort of architecture could do what Betty did?

Many explanations are compatible with any observed performance, e.g.:

- Pure chance?
- An innate behaviour triggered by some mixture of internal and external state?
  - What mixture?
  - How did the genes get the information? Why was it selected?
- A learnt adaptation in a trainable (altricial) reactive system?
  - What sort of boot-strapping could achieve this?
  - How is the learnt information acquired, represented, stored, activated, used?
- Was it a deliberative (e.g. problem-solving) process?
  - Using what sort of ontology for possible goals, states, actions?
  - Using what general knowledge?
  - Invoked how?
  - Acquired how? (Using an architecture built in infancy?)
  - Using what planning mechanisms? (Using what representations, what search mechanisms?)
- Did it involve self-knowledge? (Reflection/meta-management)
  - Did Betty understand what she was doing, or did she, like many AI deliberative systems, lack reflection/meta-management? (Can a crow teach another crow to do this?)

The questions are deep and important because understanding of spatio-temporal processes can be re-used in many contexts.

E.g. doing mathematics, designing architectures, thinking about anything complex.

Vision and affordances

Vision is not just about

- Object recognition
- Perception of geometrical and physical structure and motion
- Building cognitive maps for route-planning

There's something deeper, not yet properly characterised, which can be called perception of affordances.

The idea comes from the psychologist, J.J.Gibson (1979) *The Ecological Approach to Visual Perception*

- Affordances are not "objective" properties intrinsic to physical configurations.
- They are relational features dependent on the perceiver's
  - Common or likely goals and needs
  - Capabilities for action (physical design + software)
  - Constraints and preferences (avoid stress, injury)

What affordances did Betty need to see?

What sort of computer-based system (robot) could see them, and use them to solve similar problems?

Why do we need machines to see and understand affordances?

Example:
Think of a tutorial system teaching mechanical engineering.

- This may have to display diagrams, e.g. exploded views of parts of engines.
- Suppose there is something the learner does not understand.
- The machine may have see how to rotate the object in the diagram to make an important bit more visible
- Or it may have to know how to select a portion to move in order to show more clearly the operation of the machine.
- These are things a human tutor can do.

If you did not see clearly what Betty the crow did in the video, could the computer understand why you could not see it clearly?

E.g. because the crucial part of the apparatus was on the far side of the tube.

EXERCISE: think of examples where a machine's ability to understand affordances is crucial to doing its job effectively.

Towards a theory of affordances

Affordances in a complex scene can be construed as

- **sets of sets** of counterfactual conditionals,
- **spatially indexed:** different sets attached to different parts of objects.

How should affordances be perceived, represented, used, explained to others?

Different representations and mechanisms handle affordances in different architectural layers -
e.g. skilled behaviour is mostly reactively controlled.

Probably not using modal logics???

Can we get some hints by exploring possible neural under-pinnings for sets of conditionals stored as context-sensitive associations?

This still leaves many hard questions about short term memory, selection of what is relevant, reasoning about possibilities, control of search, etc.

The future is your responsibility

How to make progress

- Make sure you understand the nature of the task of an HCI system.

  E.g. teaching, giving information, collecting information, preparing essays, developing programs, collaborative design or problem solving, etc.
- Use shallow designs when they are adequate to the task,

  Makking sure a design is adequate can be tricky. What works for you and your friends may not work for people from another background, or with different intellectual powers.
- When a shallow design is not adequate, use
  - a theory of how human minds work when they perform the task
  - A theory of how human minds work when they interact with someone performing the task, e.g. teaching them, or asking for help.
- Try to use that to inform your design

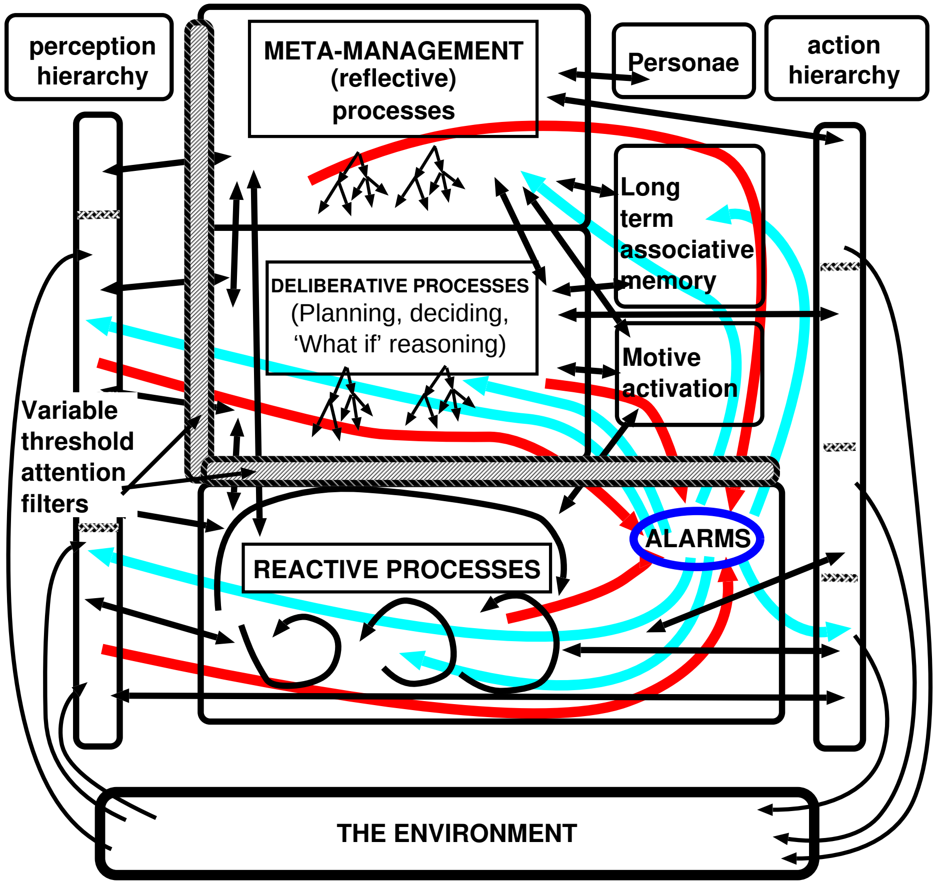
Perhaps you can change the world of interface design.

YOU WILL NEED A GOOD THEORY OF HOW HUMAN MINDS WORK
PERHAPS A THEORY DEVELOPED IN THE COGAFF PROJECT?

http://www.cs.bham.ac.uk/research/cogaff/

A hypothetical Human-like architecture:
H-CogAff (See http://www.cs.bham.ac.uk/research/cogaff/)

This partly overlaps with Minsky's *Emotion machine* architecture.



Where could it come from?

Various trajectories: evolutionary, developmental, adaptive, learning...