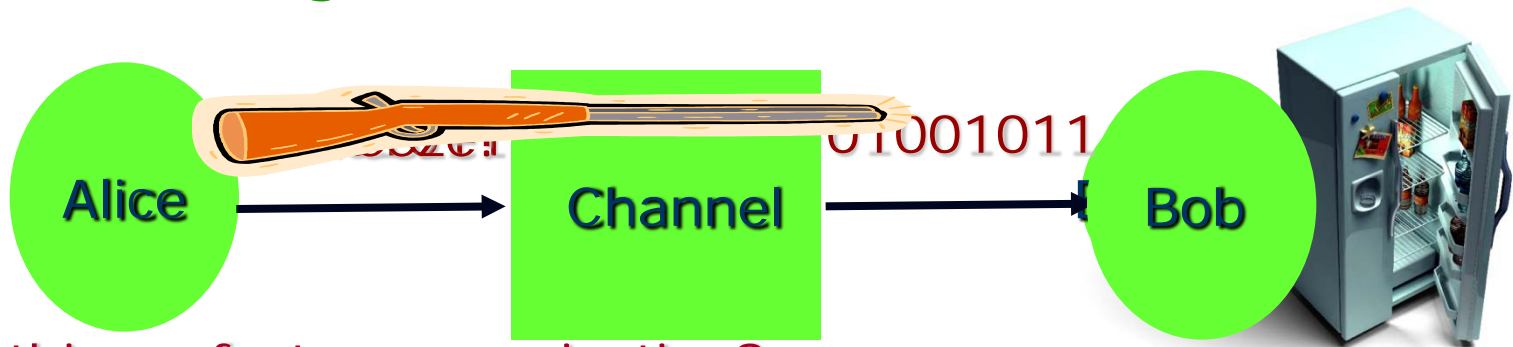# Semantic Goal-Oriented Communication

Madhu Sudan
Microsoft Research + MIT

Joint with **Oded Goldreich** (Weizmann) and **Brendan Juba** (MIT).

Disclaimer

- Work in progress (for ever) ...

- Comments/Criticisms/Collaboration/Competition welcome.

The Meaning of Bits

- Is this perfect communication?

- What if Alice is trying to send instructions?
  - Aka, an algorithm
  - Does Bob understand the correct algorithm?
  - What if Alice and Bob speak in different (programming) languages?

Miscommunication (in practice)

- Exchanging (powerpoint) slides.
  - Don't render identically on different laptops.
- Printing on new printer.
  - User needs to "learn" the new printer, even though printer is quite "intelligent".
- Many such examples ...
  - In all cases, sending bits is insufficient.
  - Notion of meaning ... intuitively clear.
  - But can it be formalized?
    - Specifically? Generically?
    - While conforming to our intuition

Bits vs. their meaning

- Say, User and Server know different programming languages. Server wishes to send an algorithm A to User.
  - A = sequence of bits … (relative to prog. language)

- Bad News: Can't be done
  - For every User, there exist algorithms A and A', and Servers S and S' such that S sending A is indistinguishable (to User) from S' sending A'

- Good News: Need not be done.
  - From Bob's perspective, if A and A' are indistinguishable, then they are equally useful to him.

- What should be communicated? Why?

Part I: Computational Motivation

Computational Goal for Bob

- Why does User want to learn algorithm?
    - Presumably to compute some function f
      (A is expected to compute this function.)
    - Lets focus on the function f.

- Setting:
    - User is prob. poly time bounded.
    - Server is computationally unbounded, does not speak same language as User, but is "helpful".
    - What kind of functions f?
        - E.g., uncomputable, PSPACE, NP, P?

Setup

User   Server

$f(x) = 0/1?$

$R \leftarrow \$\$\$$

q_1

Different from interactions in cryptography/security:

There, User does not trust Server, while here he does not understand her.

Computes $P(x, R, a_1, \ldots, a_k)$

Hopefully $P(x, \ldots) = f(x)$!

Intelligence & Cooperation?

- For User to have a non-trivial interaction, Server must be:
  - Intelligent: Capable of computing f(x).
  - Cooperative: Must communicate this to User.
- Formally:
  - Server S is f-<u>helpful</u> if
    $\exists$ some (other) user U′ s.t.
    $\forall$ x, starting states σ of the server
    (U′(x) $\leftrightarrow$ S(σ)) outputs f(x)

Successful universal communication

- Universality: Universal User U should be able to talk to any (every) f-helpful server S to compute f.

- Formally:
    - U is f-universal, if
      $$\forall \text{ f-helpful } S, \forall \sigma, \forall x$$
      $$(U(x) \leftrightarrow S(\sigma)) = f(x) \text{ (w.h.p.)}$$

- What happens if S is not helpful?
    - Paranoid view $\Rightarrow$ output "f(x)" or "?"
    - Benign view $\Rightarrow$ Don't care (everyone is helpful)

Main Theorems [Juba & S. '08]

- If f is PSPACE-complete, then there exists a f-universal user who runs in probabilistic polynomial time.
  - Extends to checkable ("compIP") problems
    - (NP ∩ co-NP, breaking cryptosystems)
    - S not helpful ⇒ output is safe

- Conversely, if there exists a f-universal user, then f is PSPACE-computable (in "compIP")
  - Scope of computation by communication is limited by misunderstanding (alone).

Proofs?

- Positive result:
    - f ∈ PSPACE ⇒ membership is verifiable.
    - User can make hypothesis about what the Server is saying, and use membership proof to be convinced answer is right, or hypothesis is wrong. Enumerate, till hypothesis is right.
- Negative result:
    - In the absence of proofs, sufficiently rich class of users allow arbitrary initial behavior, including erroneous ones.
    - (Only leads to finitely many errors …)

Implications

- Communication is not unboundedly helpful ☹
    - If it were, should have been able to solve every problem (not just (PSPACE) computable ones).
- But there is gain in communication:
    - Can solve more complex problems than on one's own, but not every such problem.
- Resolving misunderstanding? Learning Language?
    - Formally No! No such guarantee.
    - Functionally Yes! If not, how can user solve such hard problems?

Principal Criticisms

- Solution is no good.
    - Enumerating hypotheses is too slow.
        - Approach distinguishes **right**/**wrong**; does not solve search problem.
        - Search problem <u>needs</u> new definitions to allow better efficiency.

- Problem is not the right one.
    - Computation is not the goal of communication. Who wants to talk to a PSPACE-complete server?

Next part of talk

Part II: Generic Goals of Communication

Aside: Communication?

- Classical "Theory of Computing"

$$X \longrightarrow \boxed{F} \longrightarrow F(X)$$

- Issues: Time/Space on DFA? Turing machines?
- Modern theory:
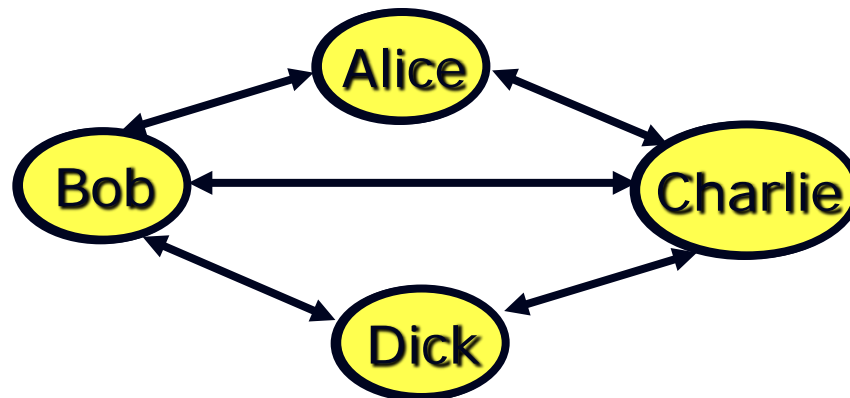
- Issues: Reliability, Security, Privacy, Agreement?
- If communication is so problematic, then why not "Just say NO!"?
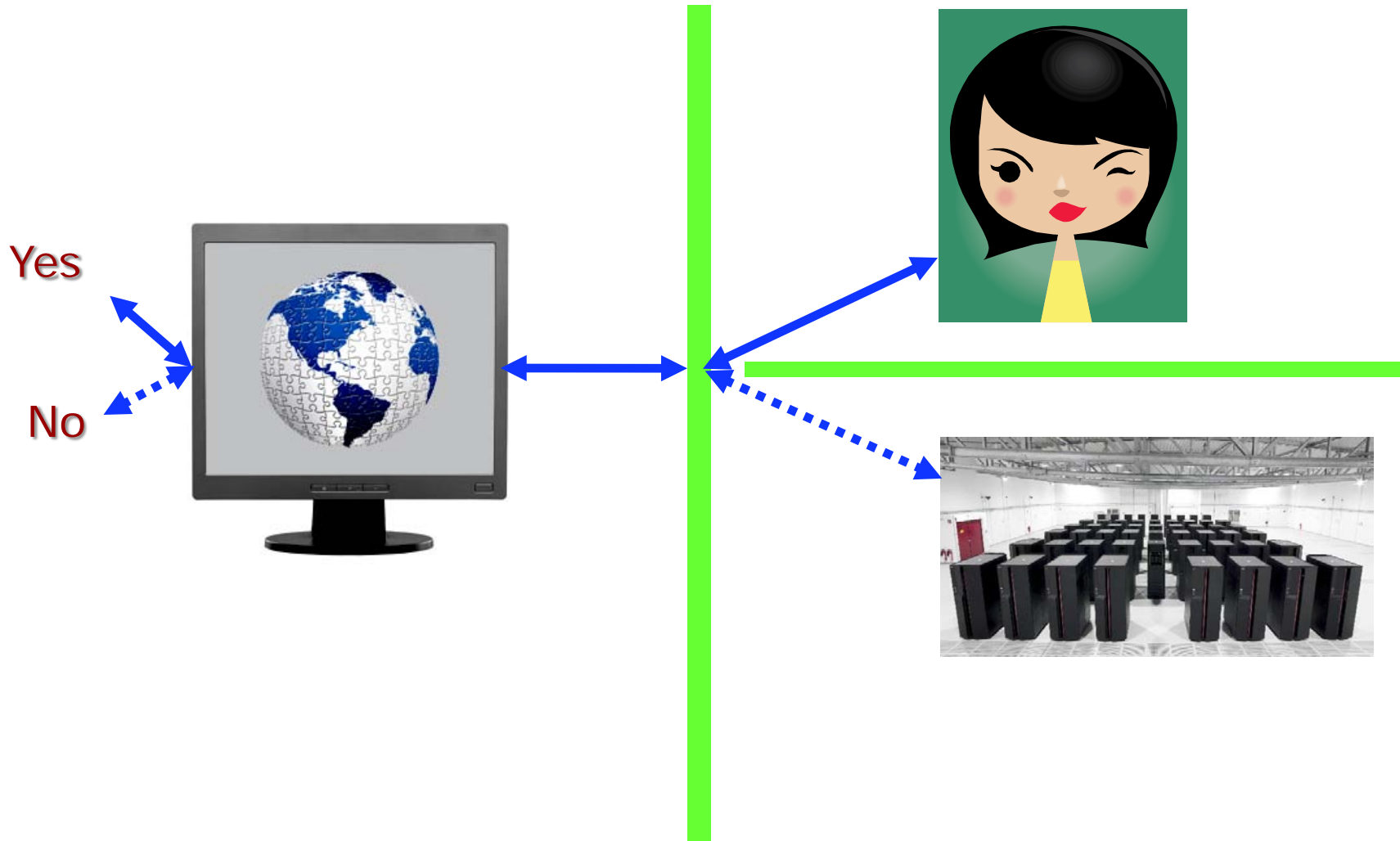
Communication: Example 1 (Printing)

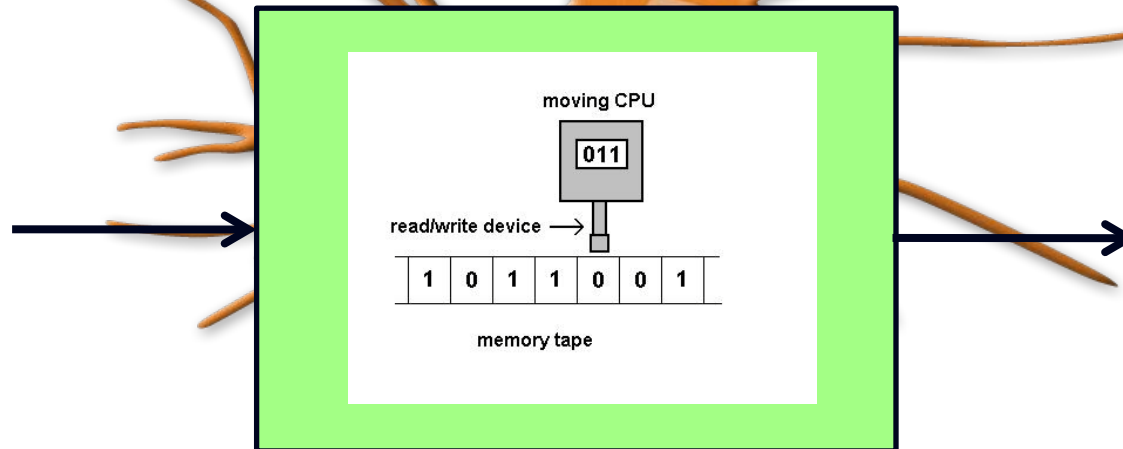# Communication: Ex. 2 (Computation)

X $f(X)$
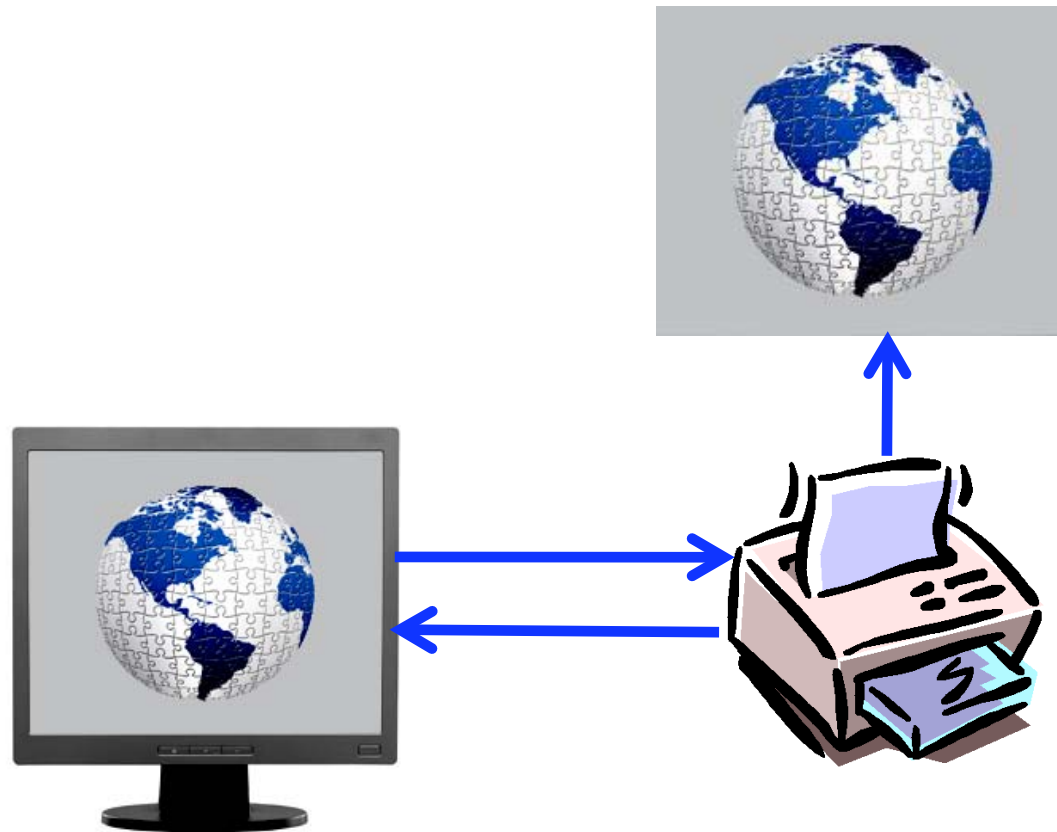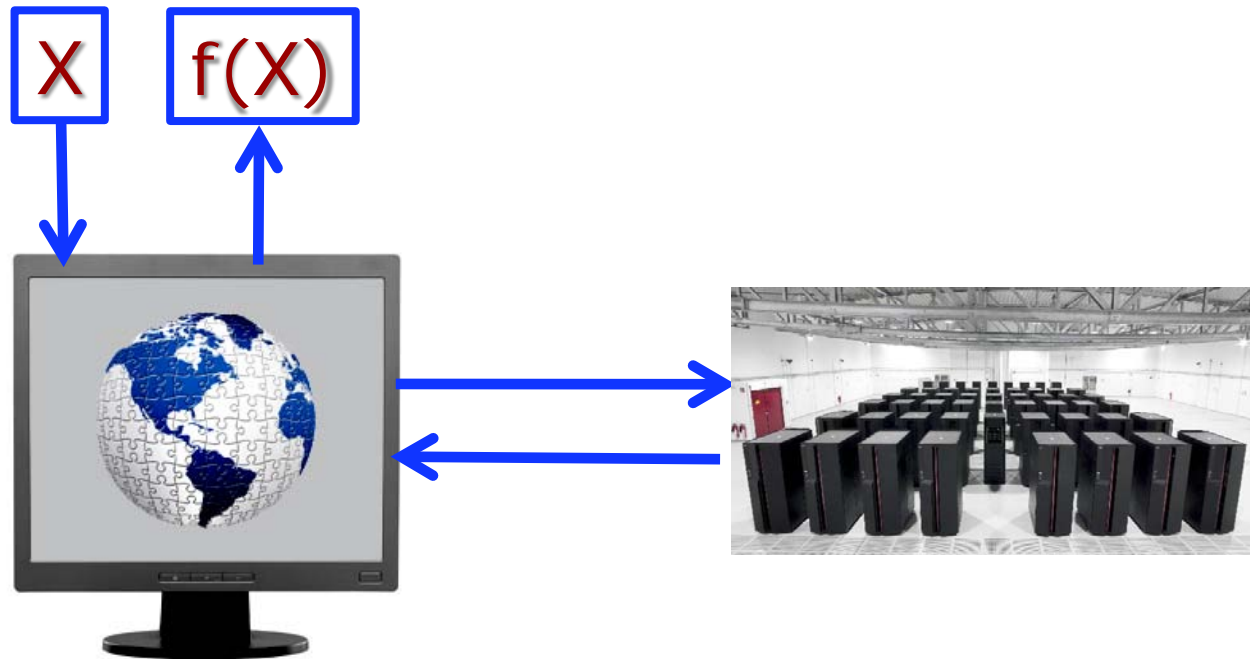
Communication: Ex. 3 (Web search)

WWW

Q

Q

P

$Q(WWW(P)) = \text{☺}?$
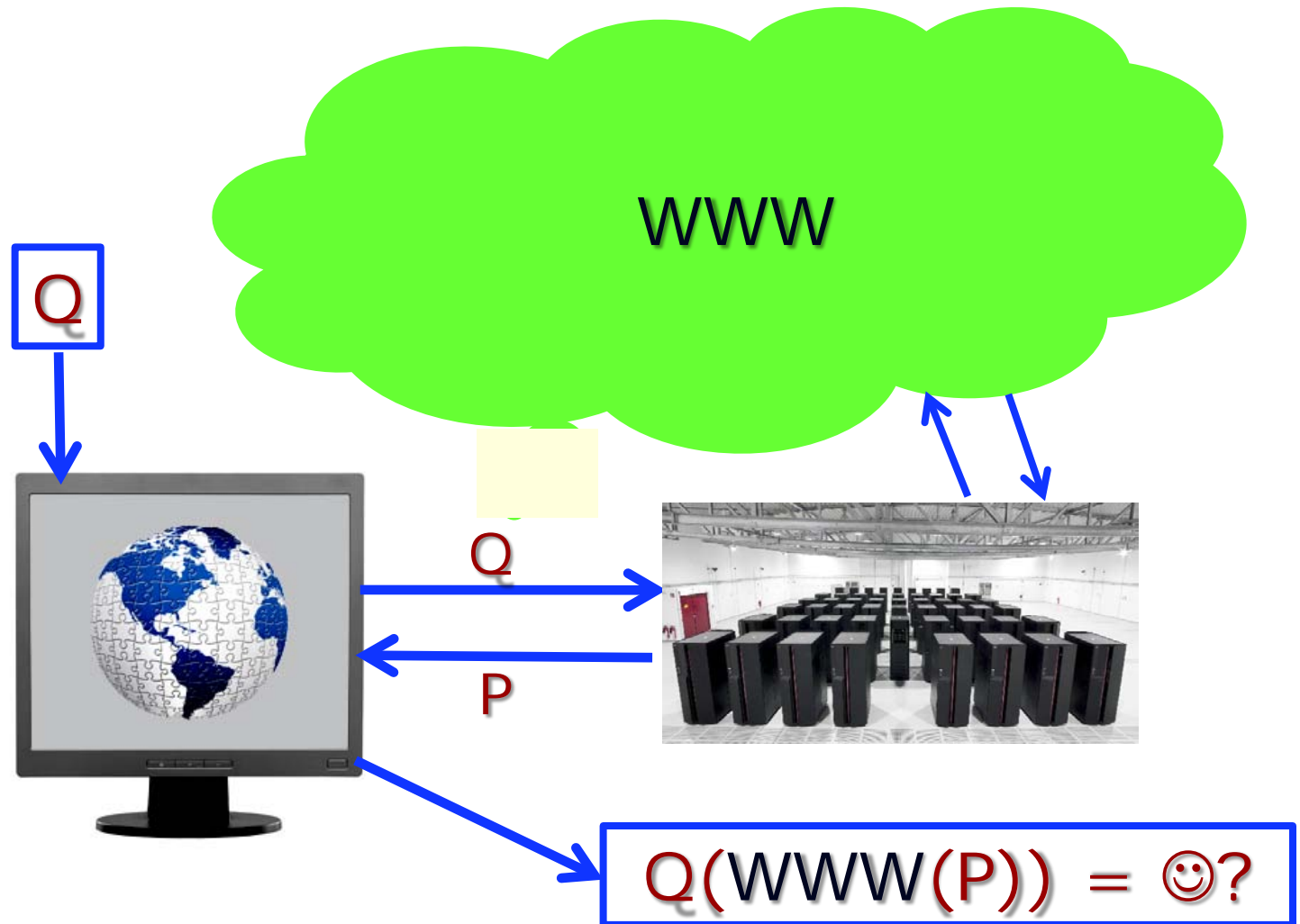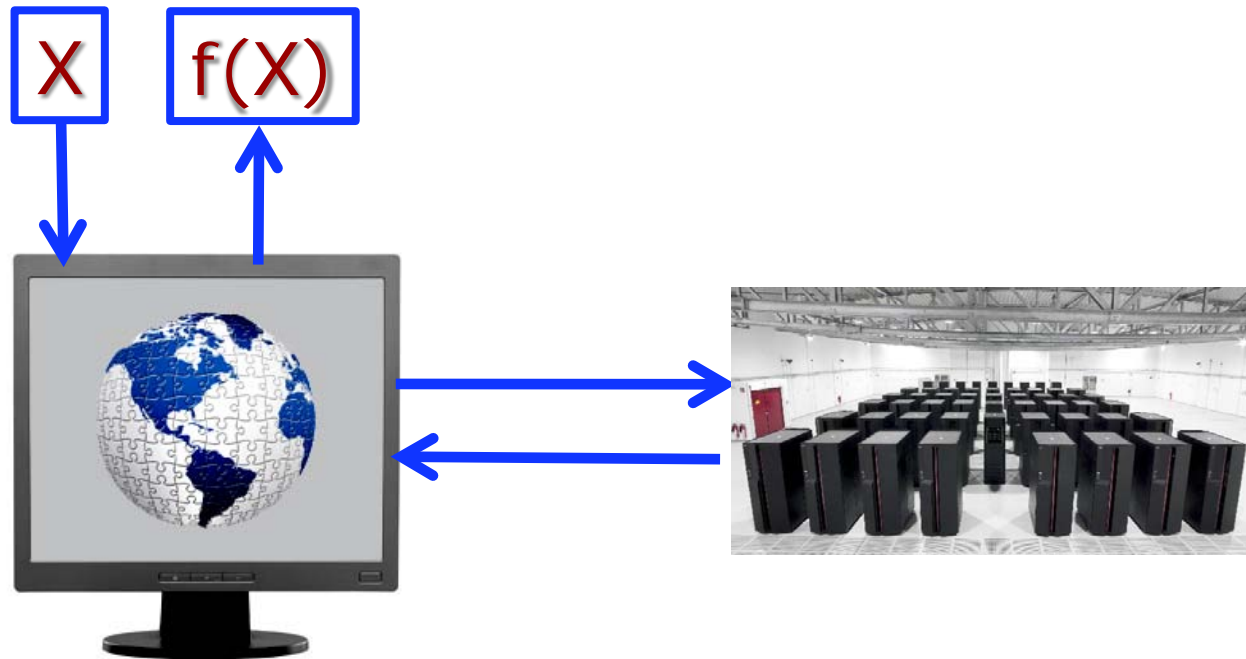
Communication: Ex. 4 (Intelligence?)

Yes

No

Aside: Modelling Computing

- Classically: Turing Machine/(von Neumann) RAM.
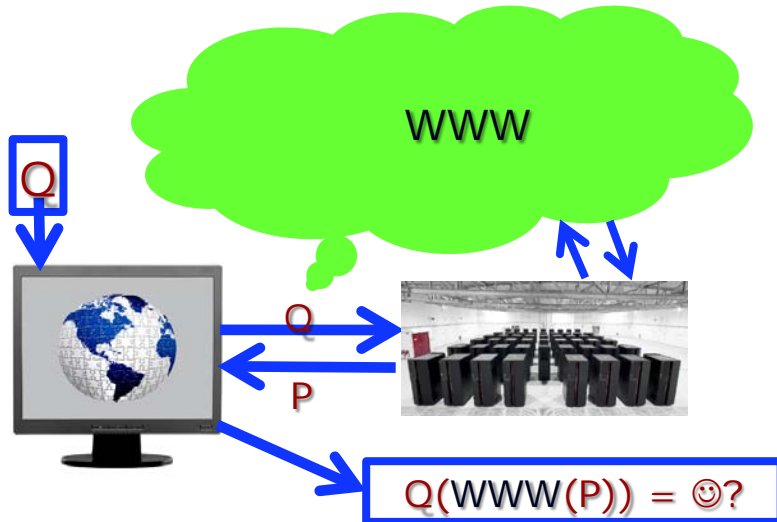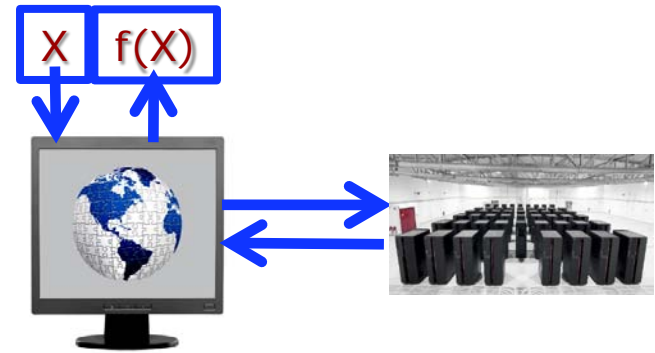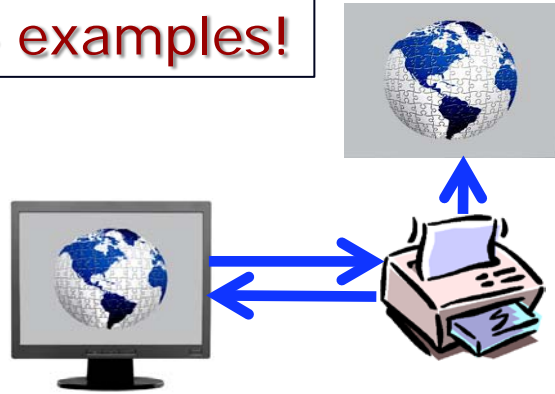 - Described most computers being built?

- Modern computers: more into communication than computing.
  - What is the mathematical model of a communicating computer?
  - What is universality?

Communication: Ex. 2 (Computation)

X $f(X)$

Generic communication problem?

Should model
All 4 examples!
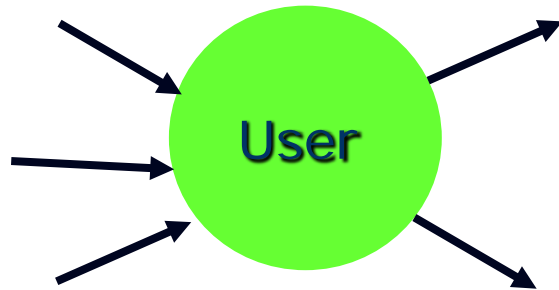
X | f(X)
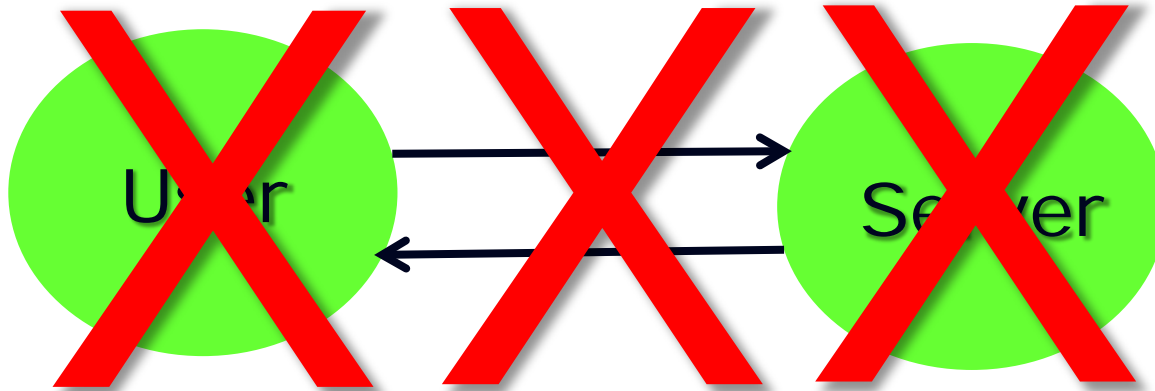
Q

WWW

Q

P

Q(WWW(P)) = ☺?

Yes
No

Modelling User/Interacting agents

- (standard AI model)

- User has state and input/output wires.
  - Defined by the map from current state and input signals to new state and output signals.
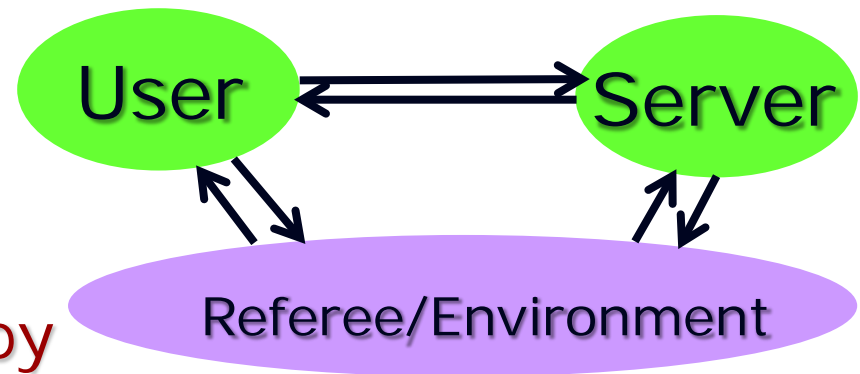
Generic Goal?

- Goal = function of ?
  - User? – But user wishes to change actions to achieve universality!
  - Server? – But server also may change behaviour to be helpful!
  - Transcript of interaction? – How do we account for the many different languages?
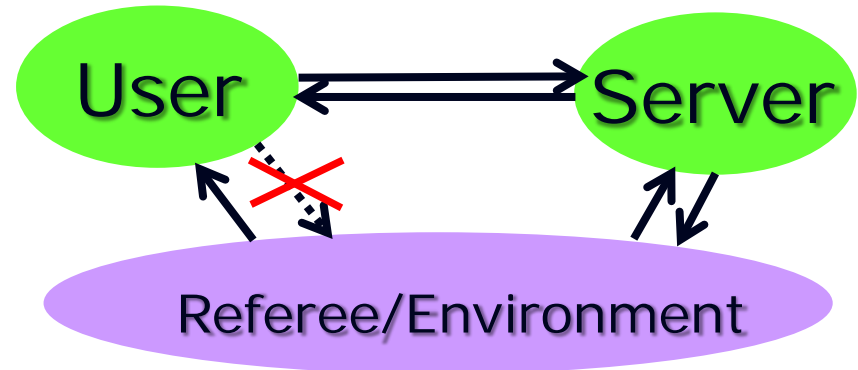
Generic Goals

- Key Idea: Introduce 3rd entity: Referee
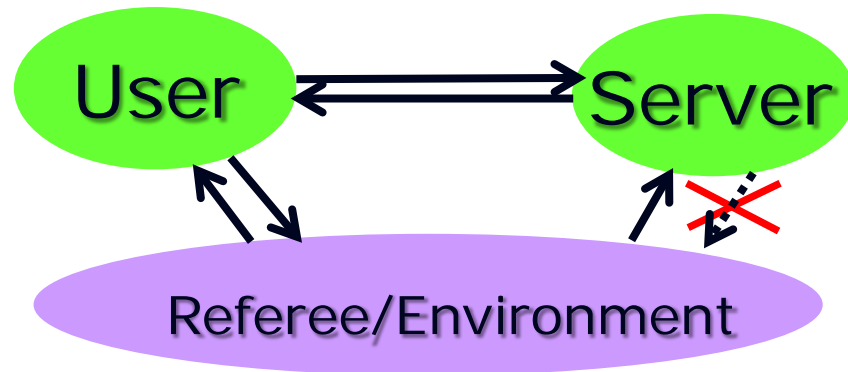 - Poses tasks to user.
  - Judges success.

- Generic Goal specified by
  - Referee (just another agent)
  - Boolean Function determining if the state evolution of the referee reflects successful achievement of goal.
  - Class of users/servers.

Generic Goals

■ Pure Control

■ Pure Informational

Sensing & Universality

- To achieve goal, User should be able to sense progress.
    - I.e., user should be compute a function that (possibly with some delay, errors) reflects achievement of goals.
- Generalization of positive result:
    - Generic goals (with technical conditions) universally achievable if ∃ sensing function.
- Generalization of negative result:
    - Sensing is necessary (in one-shot goals)
    - (In infinite goals, If non-trivial generic goal is achieved with sufficiently rich class of helpful servers, then it is safely achieved with every server.)

Conclusions

- Is there a universal communication protocol?
  - No! (All functions vs. PSPACE-computable functions).
  - But can achieve "sensible" goals universally.
  - But … diversity of goals may be the barrier to universality.

- Goals of communication.
  - Should be studied more.
  - Suggests good heuristics for protocol design:
    - Server = Helpful?
    - User = Sensing?

Language Learning

- Meaning = end effect of communication.
  - [Dewey 1920s, Wittgenstein 1950s]

- What would make learning more efficient?
  - What assumptions about "language"?
  - How to do encapsulate it as "class" restrictions on users/servers.
  - What learning procedures are efficient?

- Time to get back to meaningful conversation!

References

- Juba & S.
  - ECCC TR07-084: http://eccc.uni-trier.de/report/2007/084/

- Goldreich, Juba & S.
  - ECCC TR09-075: http://eccc.uni-trier.de/report/2009/075/

Thank You!