Semantic Goal-Oriented Communication

Madhu Sudan
Microsoft Research + MIT

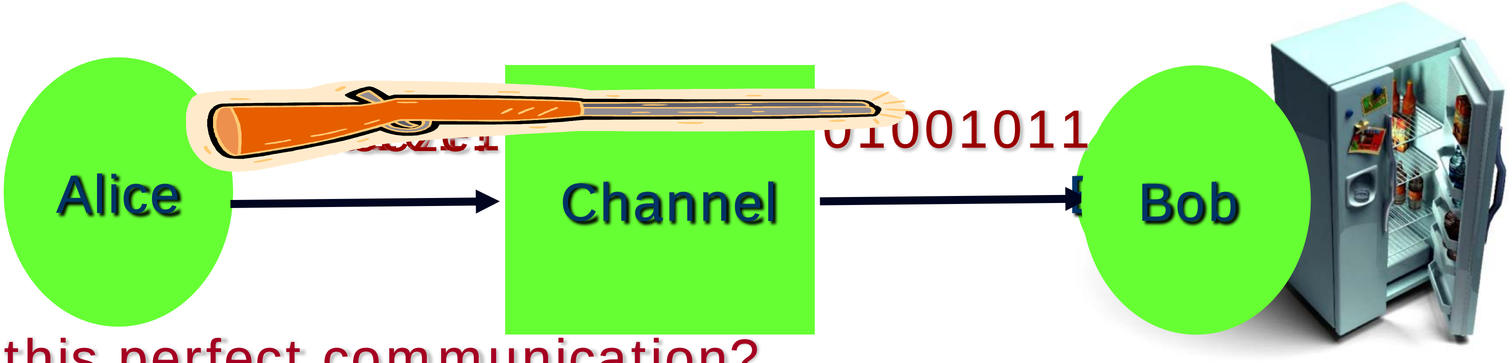
Joint with **Oded Goldreich** (Weizmann) and **Brendan Juba** (MIT).

Disclaimer

- Work in progress (for ever) …

- Comments/Criticisms/Collaboration/Competition welcome.

The Meaning of Bits



- Is this perfect communication?

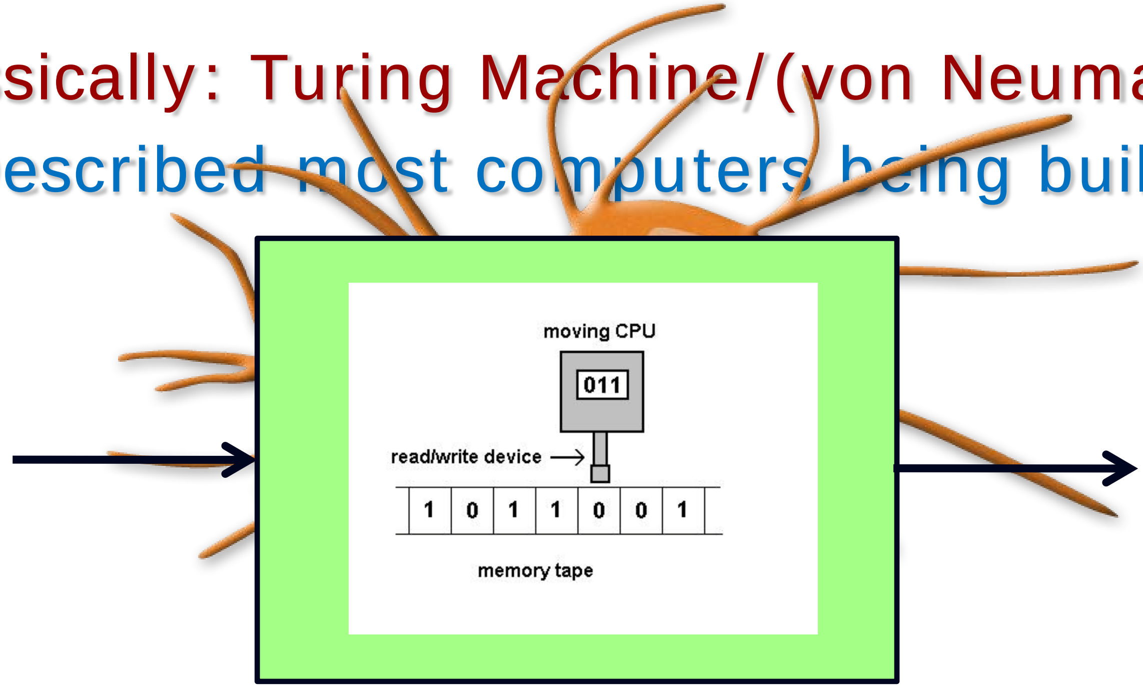
- What if Alice is trying to send instructions?
    - Aka, an algorithm
    - Does Bob understand the correct algorithm?
    - What if Alice and Bob speak in different (programming) languages?

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

Modelling Computing

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



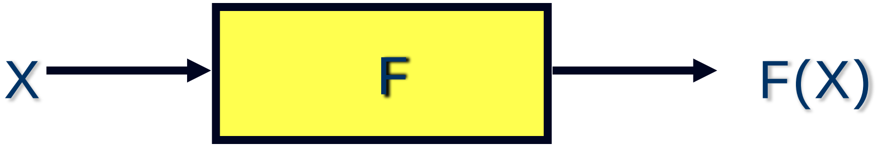
- Modern computers: more into communication than computing.
  - What is the mathematical model?
  - Do we still have universality?

Bits vs. their meaning

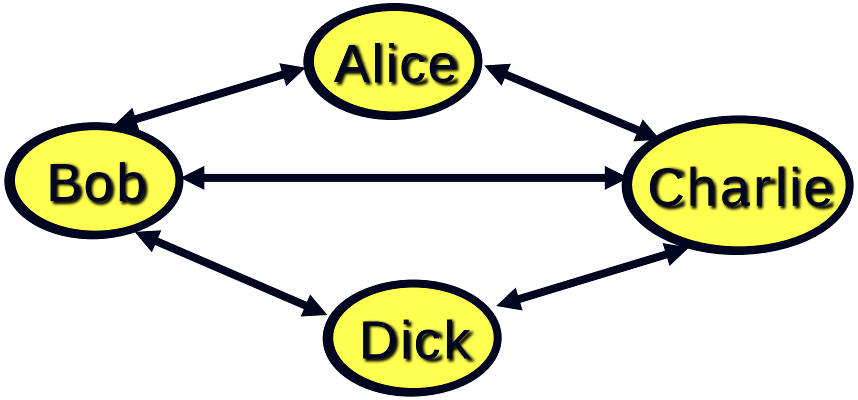
- Say, User and Server know different programming languages. Server wishes to send an algorithm A to User.
  - A = sequence of bits … (relative to prog. language)
- Bad News: Can’t be done
  - For every User, there exist algorithms A and A’, and Servers S and S’ such that S sending A is indistinguishable (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?

Aside: Why communicate?

- Classical "Theory of Computing"



- 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!"?

Motivations for Communication

- Communicating is painful. There must be some compensating gain.

- What is User's Goal?
  - "Control": Wants to alter the state of the environment.
  - "Intellectual": Wants to glean knowledge (about universe/environment).

- Thesis: By studying the goals, can enable User to overcome linguistic differences (and achieve goal).

Part II: 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?

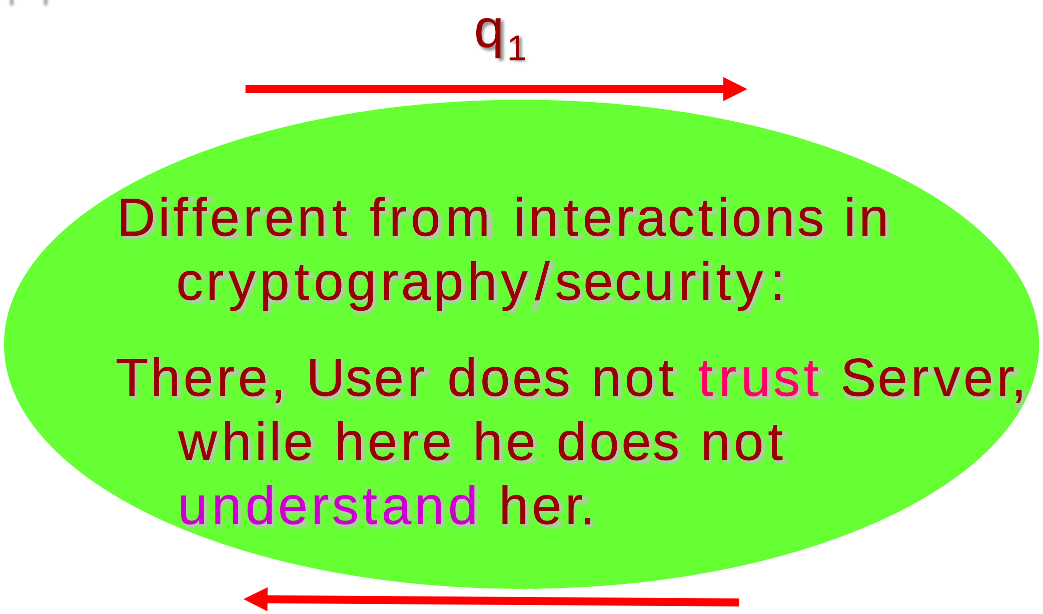
Setup

User

Server

f(x) = 0/1?

R ← $$$



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,…,a_k)

Hopefully P(x,…) = 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 helpful (for f) 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) helpful server S to compute f.

- Formally:
  - U is universal, if

    $\forall$ 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 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.
  - Scope of computation by communication is limited by misunderstanding (alone).

Proofs?

- Positive result:
  - f Є 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.
- 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

- No universal communication protocol ☹
  - If there 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 interpreters 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 III: Generic Goals

Generic Communication [Goldreich, J., S.]

- Not every goal is computational!
  - Even if it is, is Semantic Communication only possible is Server is “much better” than User?
- What are generic goals?
  - Why do we send emails?
  - Why do I browse on Amazon?
  - Why do we listen to boring lectures?
    - (or give inspirational ones ☺)
- Seemingly rich diversity

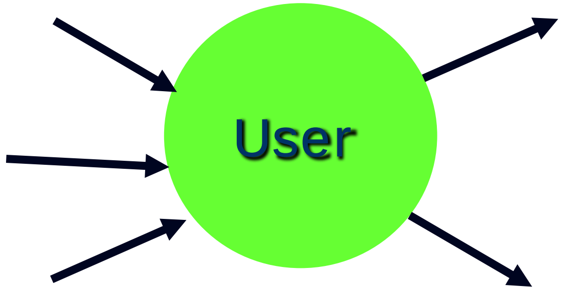
Universal Semantics for such Goals?

- Can we still achieve goal without knowing common language?
  - Seems feasible …
    - If user can detect whether goal is being achieved (or progress is being made).
  - Just need to define
    - Sensing Progress?
    - Helpful + Universal?
    - …
    - Goal?
    - User?

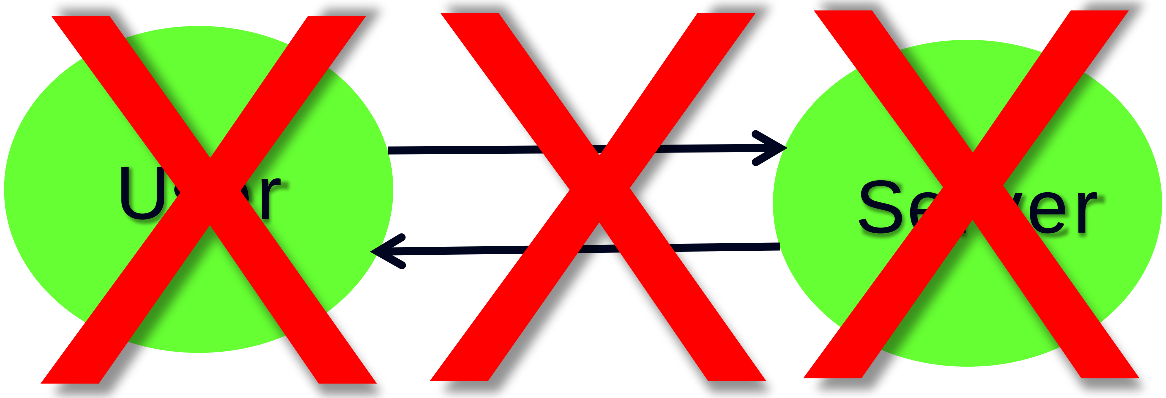
Modelling User/Interacting agents

- (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.



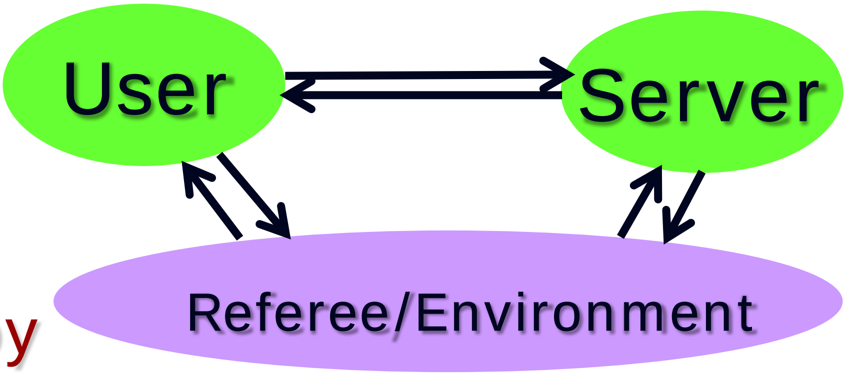
Generic Goal?



- 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

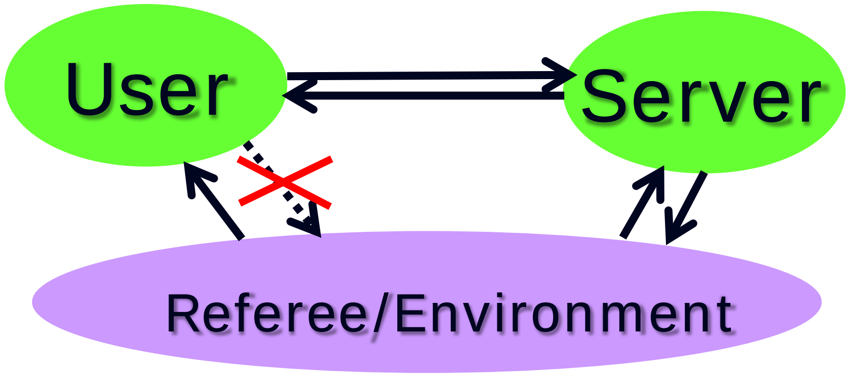
- Key Idea: Introduce 3rd entity: Referee
  - Poses tasks to user.
  - Judges success.



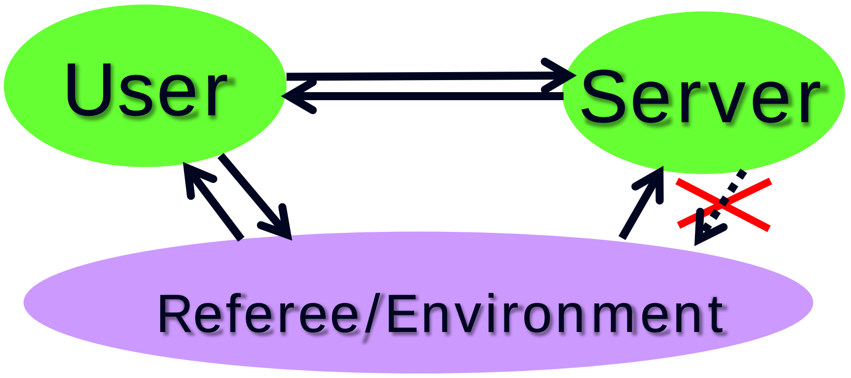
- Generic Goal specified by
  - Referee (just another agent)
  - Boolean Function determining if the state evolution of the referee reflects successful achievement of goal.
  - Class of users/servers.

Generic Goals

- Pure Control



- Pure Informational



Sensing & Universality (Theorems)

- 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.
  - “User simulates Referee”
- Generalization of positive result:
  - Generic goals (with technical conditions) universally achievable if ∃ sensing function.
- Generalization of negative result:
  - If non-trivial generic goal is achieved with sufficiently rich class of helpful servers, then it is safely achieved with every server.

Why is the paper so long?

- To capture fully general models ...
  - User/Server live for ever and Goal achievement is a function of infinite sequence of states.
  - User/Server should be allowed to be probabilistic.
  - Referee should be allowed to be non-deterministic.
  - (Getting quantifiers right non-trivial.)

When Server is less powerful than User

- What should the goal be?
  - Can’t expect server to solve (computational) problems user can’t.
  - So what can user try to do? Why?
    - Ask Server: Repeat after me …
    - Test if Server has some computational power … solves simple (linear/quadratic time) problems.
    - Has memory … can store/recall.
    - Can act (pseudo-)independently – is not deterministic, produces incompressible strings.
    - Can challenge me with puzzles.
- Each Goal/combo can be cast in our framework.
  - (Sensing functions do exist; communication is essential to achieving Goals; problems are more about control …)

(Generalized) Turing tests

- Distinguish between “Intelligent”/”Not”
- Distinguish between “Humans”/”Bots”
    - Generically:
        - Class of servers = H union N:
            - H = { (1,i) | i }
            - N = { (0,i) | i }
            - Referee: gets identity of server from server (b,s),; and after interaction between user and server, gets User’s guess b’. Accepts if b = b’.

- Sensing function exists? Depends on H vs. N.

Conclusions - 1

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

References

- Juba & S. (computational)
  - ECCC TR07-084: http://eccc.uni-trier.de/report/2007/084/
- Goldreich, Juba & S. (generic)
  - ECCC TR09-075: http://eccc.uni-trier.de/report/2009/075/
- Juba & S. – 2. (examples)
  - ECCC TR08-095: http://eccc.uni-trier.de/report/2008/095/

Thank You!