Universal *Semantic* Communication

Madhu Sudan
MIT CSAIL

Joint work with Brendan Juba (MIT).

An fantasy setting (SETI)

Alice

01001 … 00

No common language!
Is meaningful
communication possible?

Bob

What should Bob's response be?

If there are further messages, are they reacting to him?

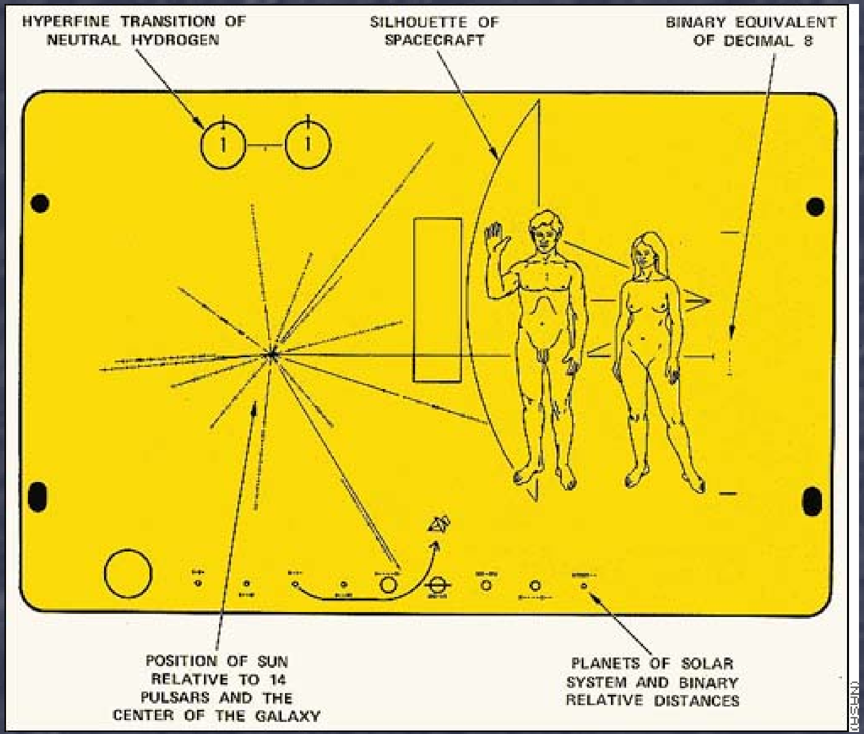
Is there an intelligent Alien (Alice) out there?

Voyager's face plate

Why did they put this image?

What would you put?

What are the assumptions and implications?



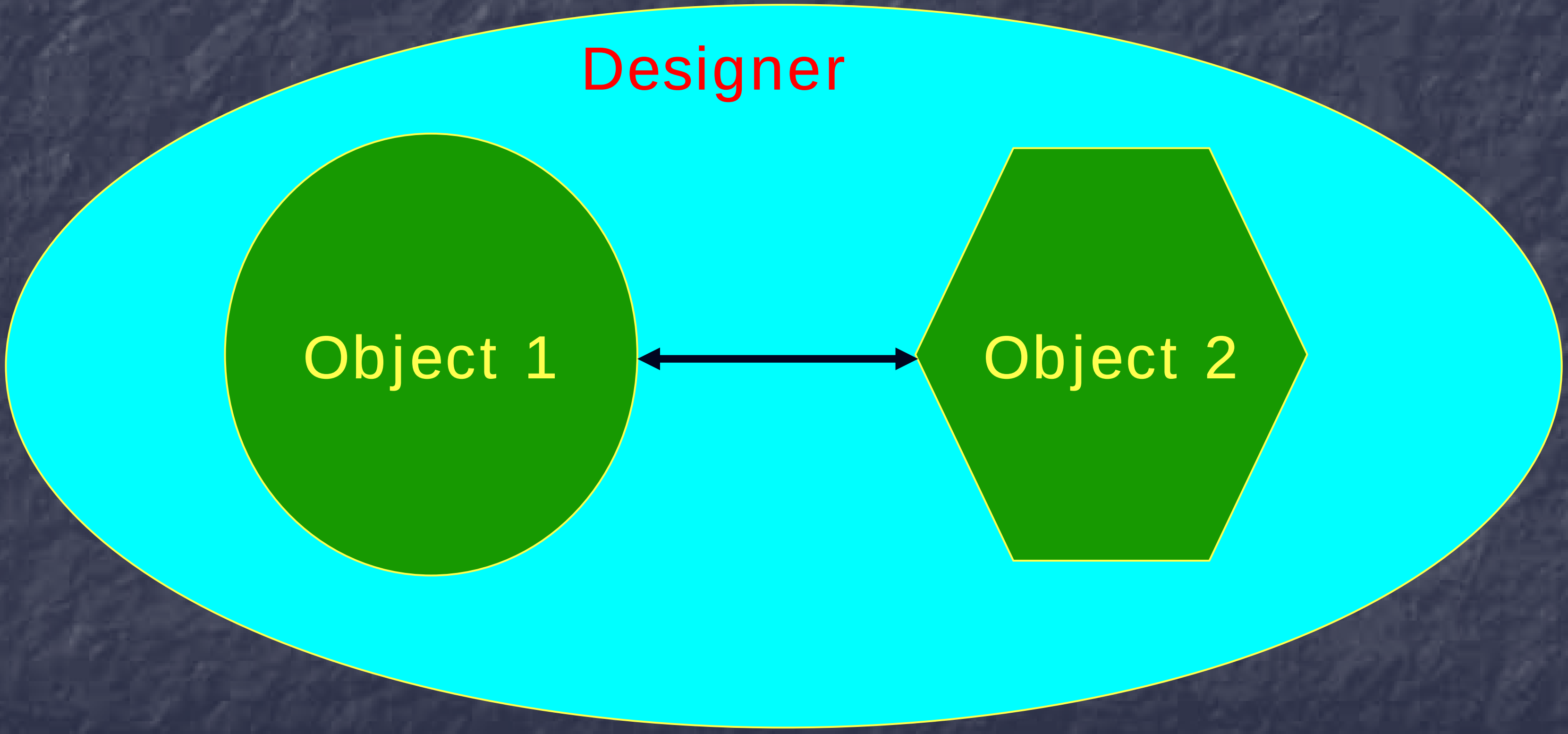
Motivation: Better Computing

- Networked computers use common languages:
  - Interaction between computers (getting your computer onto internet).
  - Interaction between pieces of software.
  - Interaction between software, data and devices.

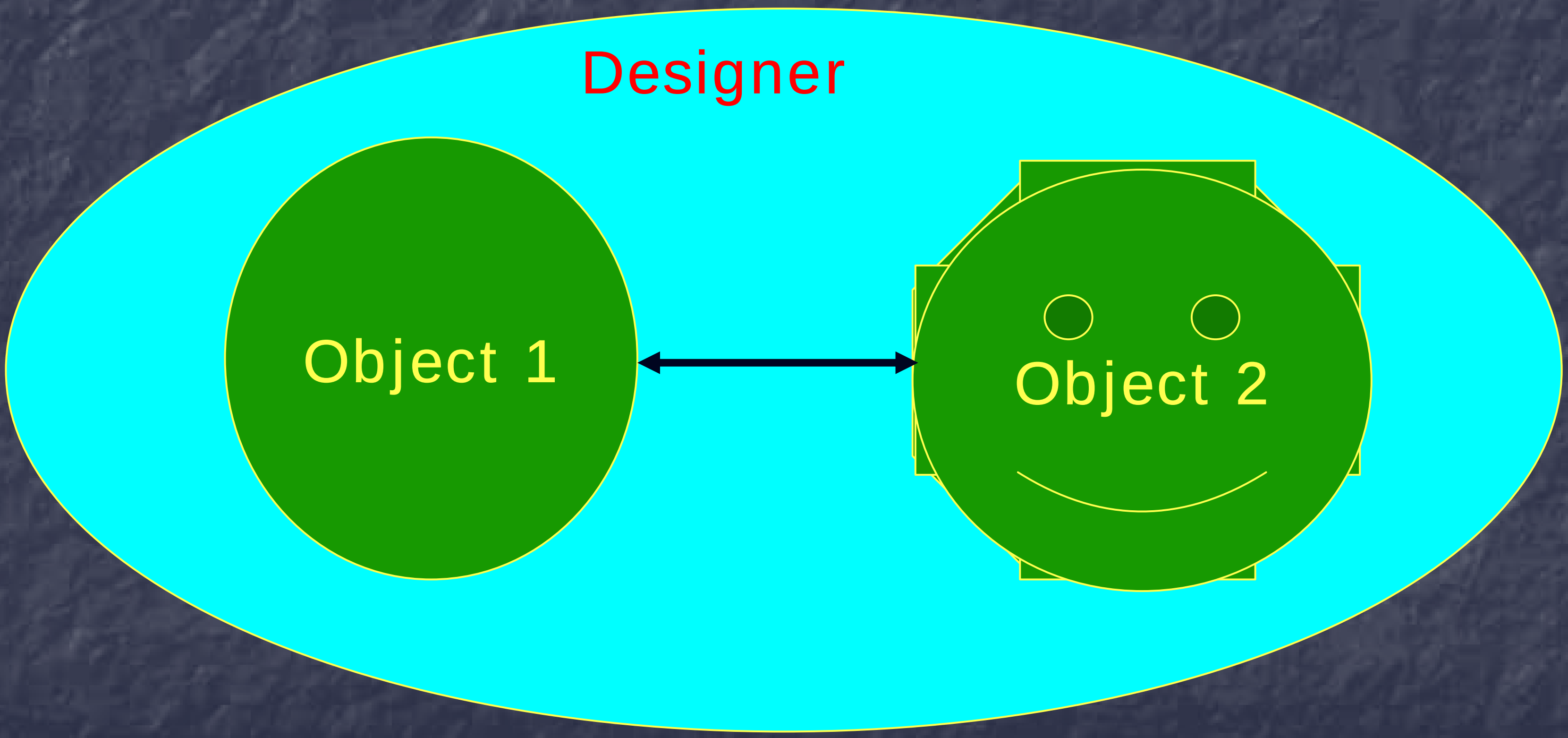
- Getting two computing environments to “talk” to each other is getting problematic:
  - time consuming, unreliable, insecure.

- Can we do this more like humans do?

Classical Paradigm for interaction



New paradigm



Robust interfaces

- Want one interface for all “Object 2”s.
- Can such an interface exist?
- What properties should such an interface exhibit?
- Puts us back in the “Alice and Bob” setting.

Goal of this talk

- Definitional issues and a definition:
  - What is successful communication?
  - What is intelligence? cooperation?
- Theorem: "If Alice and Bob are intelligent and cooperative, then communication is feasible" (in one setting)
- Proof ideas:
  - Suggest:
    - Protocols, Phenomena …
    - Methods for proving/verifying intelligence

A first attempt at a definition

- Alice and Bob are “universal computers” (aka programming languages)
- Have no idea what the other’s language is!
- Can they learn each other’s language?

- Good News: Language learning is finite. Can enumerate to find translator.

- Bad News: No third party to give finite string!
  - Enumerate? Can’t tell right/wrong ☹

Communication & Goals

- Indistinguishability of Right/Wrong: Consequence of “communication without goal”.
- Communication (with/without common language) ought to have a “Goal”.
- Bob’s Goal:
  - Verifiable: Easily computable function of interaction;
  - Complete: Achievable with common language.
  - Non-trivial: Not achievable without Alice.

Part I: A Computational Goal

Computational Goal for Bob

- Bob wants to solve hard computational problem:
  - Decide membership of L.

- Can Alice help him?

- What kind of sets L?
  - E.g., L = {set of programs P that are not viruses}.
  - L = {non-spam email}
  - L = {winning configurations in Chess}
  - L = {(A,B) | A has a factor less than B}

Setup

Which class of sets?

Bob

Alice

$x \in L?$

$R \leftarrow \$\$\$$

q_1 →

← a_1

•

•

•

q_k →

← a_k

$$G(x, R, a_1, \ldots, a_k) = 1?$$

Intelligence & Cooperation?

- For Bob to have a non-trivial exchange Alice must be
  - Intelligent: Capable of deciding if x in L.
  - Cooperative: Must communicate this to Bob.

- Formally:

Alice is L-helpful

if $\exists$ probabilistic poly time (ppt) Bob B' s.t.
$A \leftrightarrow B'(x)$ accept w.h.p. iff $x \in L$.
(independent of the history)

Successful universal communication

- Bob should be able to talk to any L-helpful Alice and decide L.

- Formally,

  Ppt B is L-universal if for every $x \in \{0,1\}^*$

  – $x \in L$ and A is L-helpful $\Rightarrow (A \leftrightarrow B(x)) = 1$ (whp).

  – $(A \leftrightarrow B(x)) = 1$ whp $\Rightarrow x \in L$.

Main Theorem

- \- If L is PSPACE-complete,
  then there exists an L universal Bob.
  (Generalizes to many other languages in PSPACE.)
- \- If there exists an L universal Bob
  then L is in PSPACE.
- In English:
  - If L is moderately stronger than what Bob can do on his own, then attempting to decide L leads to non-trivial conversation.
  - If L too strong, then leads to ambiguity.
  - Uses IP=PSPACE [LFKN,Shamir]

Contrast with Interactive Proofs

- Similarity: Interaction between Alice and Bob.
- Difference: Bob does not trust Alice.

  (In our case Bob does not understand Alice).

- Famed (hard) theorem: IP = PSPACE.
  - Membership in L can be proved interactively to a ppt. Bob.
  - Needs a PSPACE-complete prover Alice.

Few words about the proof

- Positive result: Enumeration + Interactive Proofs
  - Bob: Verifies $x \in L$ by simulating IP verifier.
  - But needs to ask the IP Prover many questions
  - Translates into many other questions $y \in L$
  - To get answers: Bob guesses Bob$'$
    - Simulates interaction between Alice and Bob$'$.

If $x \in L$ and Bob$'$ is correct, get a convincing proof.

If proof is convincing $x \in L$!

Few words about the proof

- Positive result: Enumeration + Interactive Proofs
- Negative result:
  - Suppose Alice answers every question so as to minimize the conversation length.
    - Reasonable(?) misunderstanding.
  - Conversation comes to end quickly.
  - Bob has to decide.
  - Decision can be computed in PSPACE (since Alice's strategy can be computed in PSPACE).
  - Bob must be wrong if L is not in PSPACE.
  - Warning: Only leads to finitely many mistakes.

Is this language learning?

- End result promises no language learning: Merely that Bob solves his problem.
- In the process, however, Bob learns Bob’
- But this may not be the right Bob’!
- All this is Good!
  - Should not attempt to distinguish indistinguishables!

Part II: Other Goals?

Goals of Communication

- Largely unexplored (at least explicitly)!

- Main categories
  - Remote Control:
    - Laptop wants to print on printer!
    - Buy something on Amazon
  - Intellectual Curiosity:
    - Listening to music, watching movies
    - Coming to this talk
    - Searching for alien intelligence

Extending results to other goals

- Generic Goal (for Bob): efficiently computable predicate of
  - Private input, randomness
  - Interaction with Alice
  - Environment (Altered by actions of Alice)

- Verifiability and non-triviality of goal (should) imply universal communication.

- Models situations of control.

How to model curiosity?

- How can Alice create non-trivial conversations? (when she is not more powerful than Bob)
  - Non-triviality of conversation depends on the ability to jointly solve a problem that Bob could not solve on his own.
  - But now Alice can't help either!
  - We are stuck?

Cryptography to the rescue

- Alice can generate hard problems to solve, while knowing the answer.
  - E.g. “I can factor N”;
  - Later “P * Q = N”
- If B’ is intellectually curious, then he can try to factor N first on his own … he will (presumably) fail. Then Alice’s second sentence will be a “revelation” …
- Non-triviality: Bob verified that none of the algorithms known to him, convert *his knowledge* into factors of N.

More generally

- Alice can send Bob a Goal function.
- Bob can try to find conversations satisfying the Goal.
- If he fails (once he fails), Alice can produce conversations that satisfy the Goal.
- Universal?

Part III: Voyager Faceplate?
Non-interactive proofs of intelligence?

Compression is universal

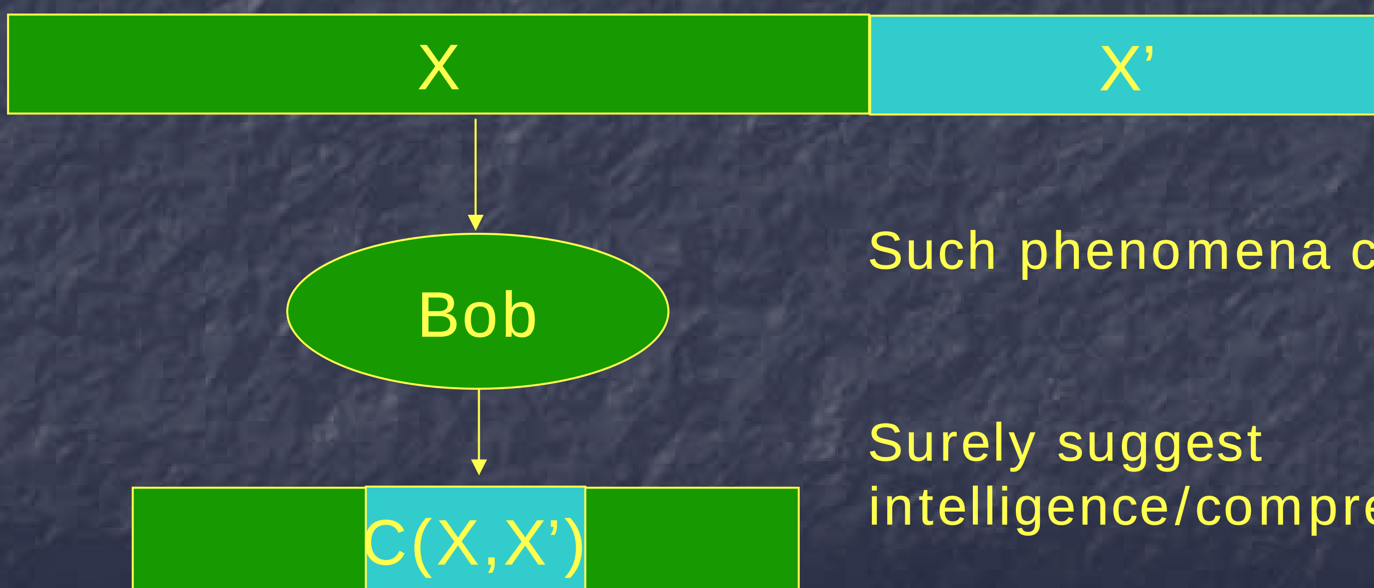
- When Bob receives Alice's string, he should try to look for a pattern (or "compress" the string).

- Universal efficient compression algorithm:
  - Input(X);
  - Enumerate efficient pairs (C(), D());
  - If D(C(X)) ≠ X then pair is invalid.
  - Among valid pairs, output the pair with smallest |C(X)|.

Compression-based Communication

- As Alice sends her string to Bob, Bob tries to compress it.

- After n steps .9n steps



Such phenomena can occur!

Surely suggest intelligence/comprehension?

Summary

- Communication should strive to satisfy one's goals.
- If one does this "understanding" follows.
- Can enable understanding by dialog:
  - Laptop -> Printer: Print <file>
  - Printer: But first tell me
    - "If there are three oranges and you take away two, how many will you have?"
  - Laptop: One!
  - Printer: Sorry, we don't understand each other!
  - Laptop: Oh wait, I got it, the answer is "Two".
  - Printer: All right … printing.

Further work

- Criticism of computational setting:
  - PSPACE Alice?
  - Exponential time learning (enumerating Bob').
    - Necessary in our model.

- What are the goals of communication?
- What are assumptions needed to make language learning efficient?

http://theory.csail.mit.edu/~madhu/papers/juba.pdf

Thank You!