# Universal Semantic Communication

**Madhu Sudan**
MIT CSAIL

Joint work with **Brendan Juba** (MIT CSAIL).

# An fantasy setting (SETI)

Alice

01001010101000111100100 0

No common language!
Is meaningful
communication possible?

Bob

What should Bob's response be?

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

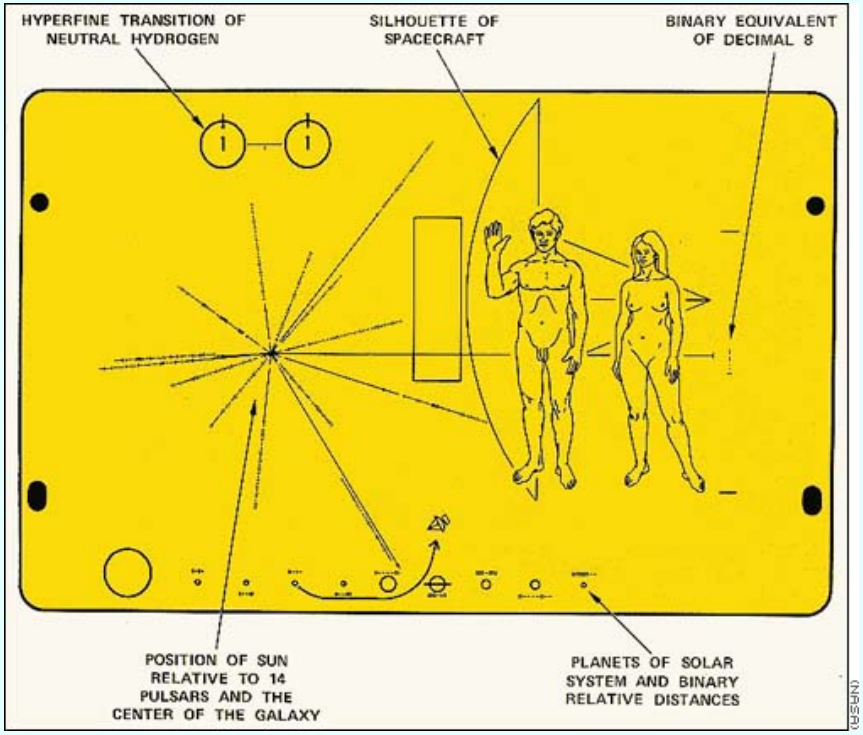
Is there an intelligent Alien (Alice) out there?

# Pioneer's face plate

Why did they put this image?

What would you put?

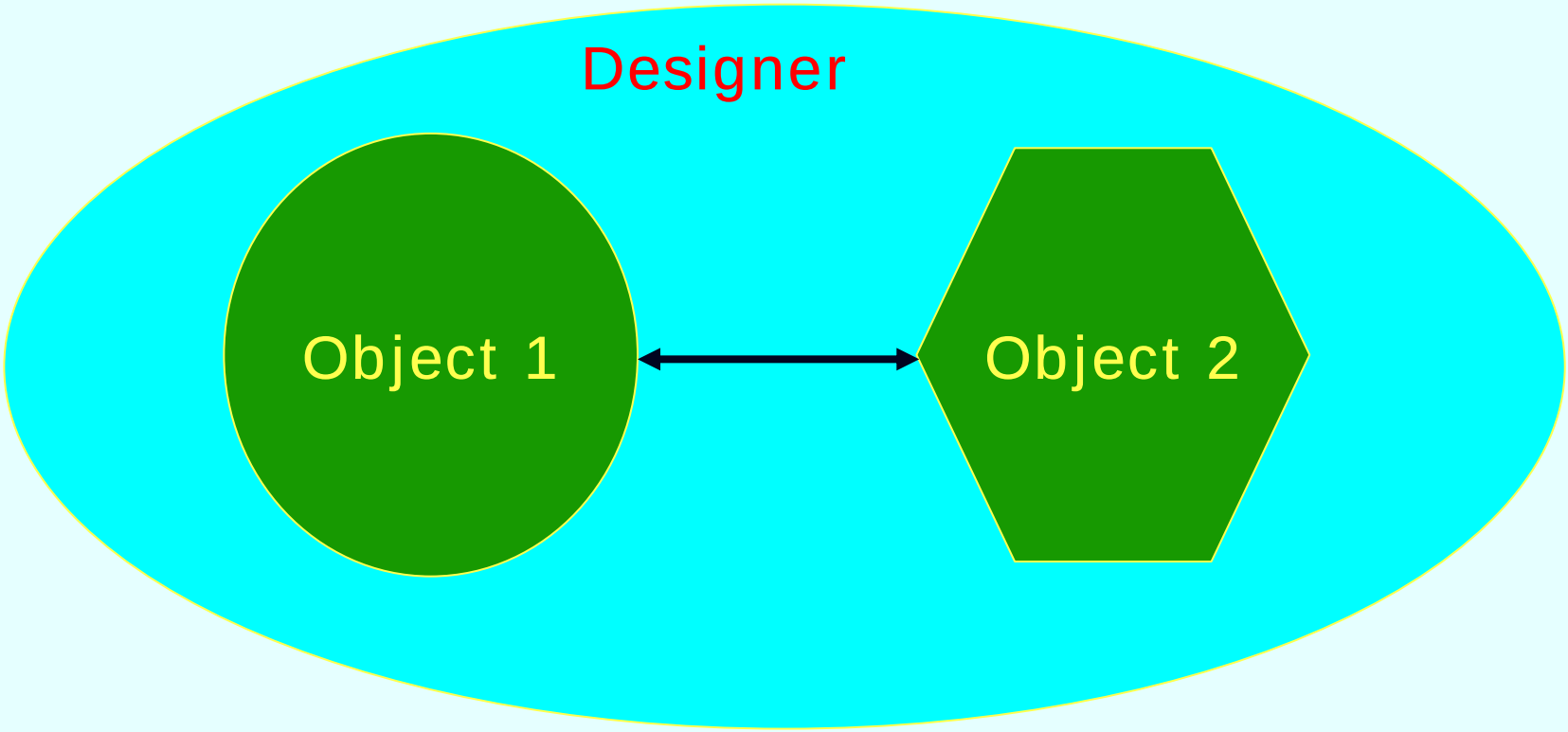
What are the assumptions and implications?



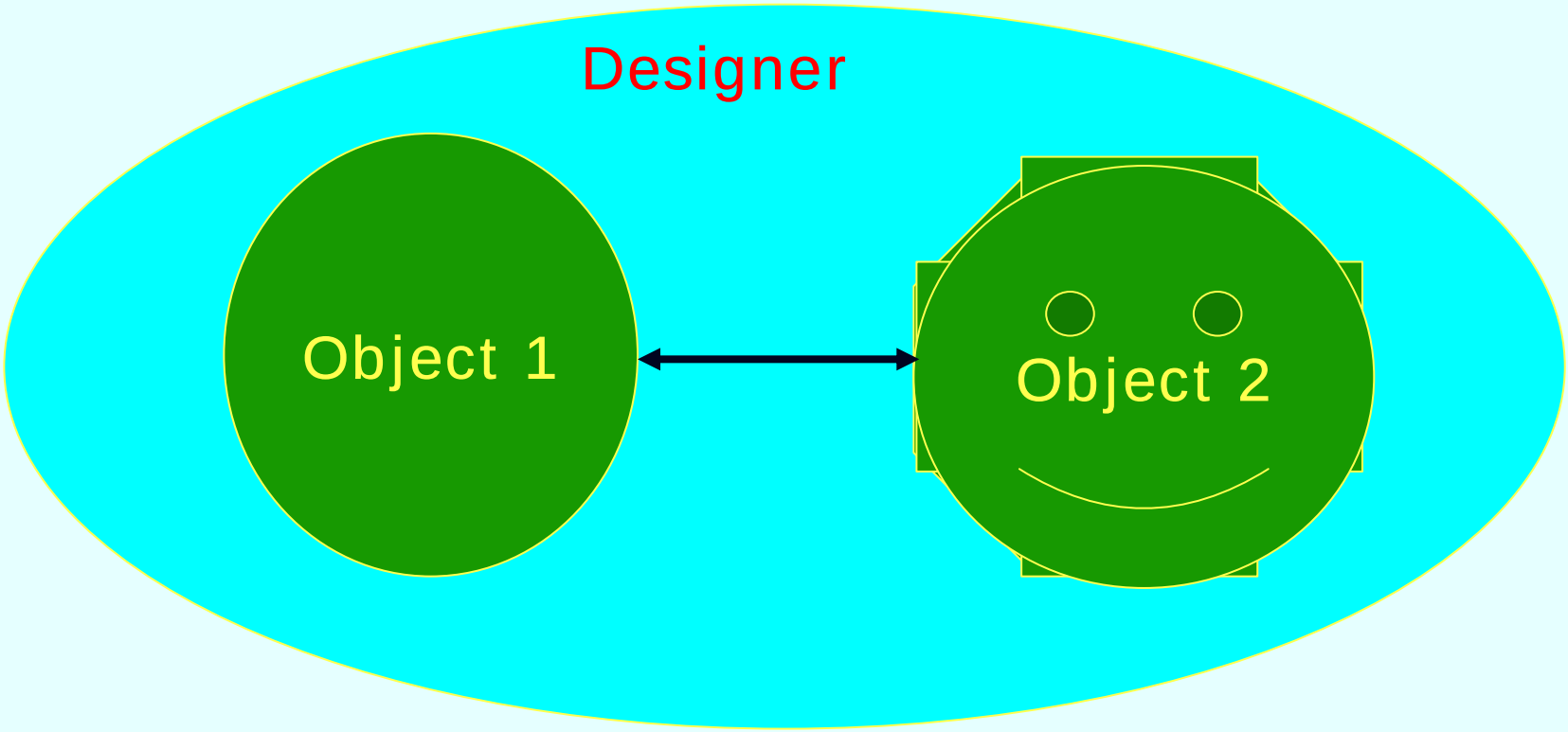
# Motivation: Better Computing

- 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 communicate 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

# What has this to do with computation?

- **In general**: Subtle issues related to "human" intelligence/interaction are within scope of computational complexity. E.g.,
  - Proofs?
  - Easy vs. Hard?
  - (Pseudo)Random?
  - Secrecy?
  - Knowledge?
  - Trust?
  - Privacy?

- **This talk**: What is "understanding"?

# 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".
- Before we ask how to improve communication, we should ask why we communicate?

*"Communication is not an end in itself,*
*but a means to achieving a Goal"*

# Part I: A Computational Goal

# Computational Goal for Bob

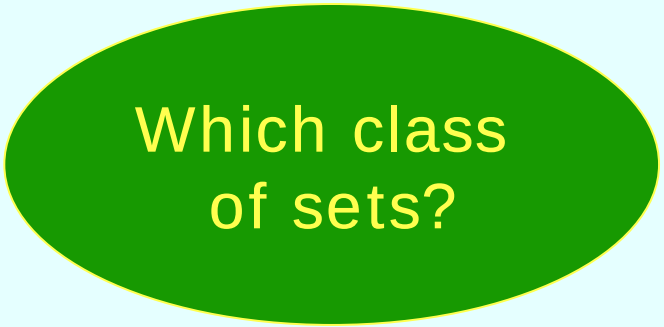
- Bob wants to solve hard computational problem:
  - Decide membership in set $S$.
- Can Alice help him?
- What kind of sets $S$? E.g.,
  - $S$ = {set of programs P that are not viruses}.
  - $S$ = {non-spam email}
  - $S$ = {winning configurations in Chess}
  - $S$ = {(A,B) | A has a factor less than B}

## Review of Complexity Classes

- **P (BPP)** – Solvable in (randomized) polynomial time (Bob can solve this without Alice's help).
- **NP** – Problems where solutions can be verified in polynomial time (contains factoring).
- **PSPACE** – Problems solvable in polynomial space (quite infeasible for Bob to solve on his own).
- **Computable** – Problems solvable in finite time. (Includes all the above.)
- **Uncomputable** (Virus detection. Spam filtering.)

Which problems can you solve
with (alien) help?

## Setup

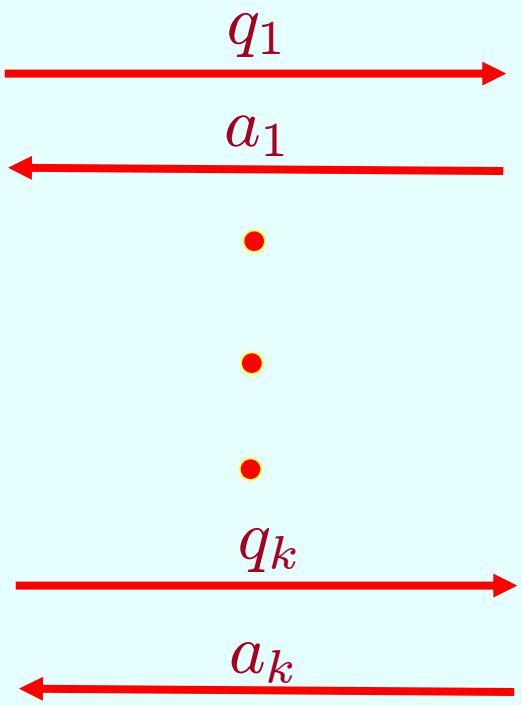



Alice

Bob

$x \in S?$

$R \leftarrow \$\$\$$




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

Hopefully $x \in S \Leftrightarrow f(\cdots) = 1$

## Contrast with Interactive Proofs

- Similarity: Interaction between Alice and Bob.
- Difference: In IP, Bob does not trust Alice.
  (In our case Bob does not understand Alice).

- Famed Theorem: IP = PSPACE [LFKN, Shamir].
  - Membership in PSPACE solvable S can be proved interactively to a probabilistic Bob.
  - Needs a PSPACE-complete prover Alice.

## Intelligence & Cooperation?

- For Bob to have a non-trivial interaction, Alice must be:
  - Intelligent: Capable of deciding if x in S.
  - Cooperative: Must communicate this to Bob.
- Modelling Alice: Maps "(state of mind,external input)" to "(new state of mind, output)".
- Formally:

Alice is S-helpful

if $\exists$ probabilistic poly time (ppt) Bob $B'$ s.t.

$\forall$ initial state of mind $\sigma$,

$A(\sigma) \leftrightarrow B'(x)$ accept w.h.p. iff $x \in S$.

## Successful universal communication

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

- **Formally,**

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

  – $A$ is $S$-helpful $\Rightarrow$ $[A \leftrightarrow B(x)] = 1$ iff $x \in S$ (whp).

  $A$ is not $S$-helpful $\Rightarrow$ Nothing!!

**Or should it be ...**

$A$ is not $S$-helpful $\Rightarrow$ $[A \leftrightarrow B(x)] = 1$ implies $x \in S$.

## Main Theorem

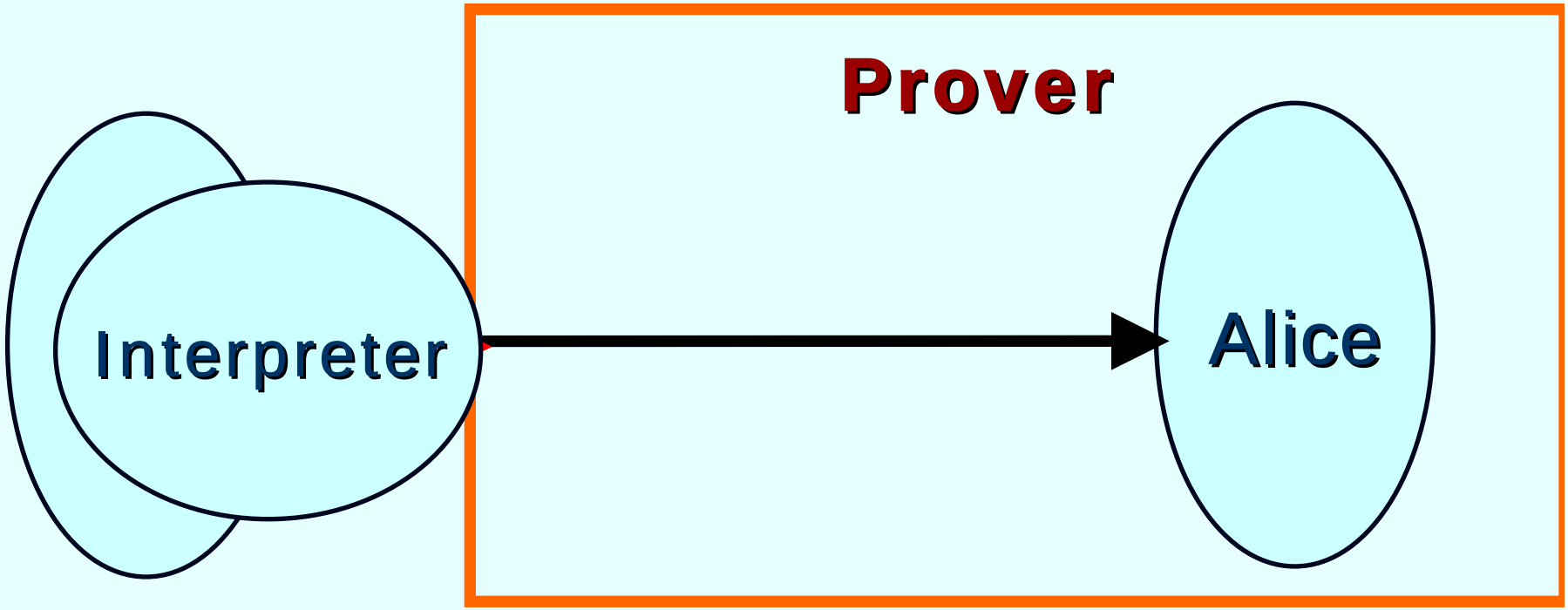
- If $S$ is PSPACE-complete (aka Chess),
  then there exists an $S$-universal Bob.
  (Generalizes to any checkable set $S$.)
- If there exists an $S$-universal Bob
  then $S$ is in PSPACE.

- **In English:**
  - **If S is moderately stronger than what Bob can do on his own, then attempting to solve S leads to non-trivial (useful) conversation.**
  - **If S too strong, then leads to ambiguity.**
  - **Uses IP=PSPACE**

# Few words about the proof

- **Positive result: Enumeration + Interactive Proofs**

Guess: Interpreter; $x \in S$?




Proof works $\Rightarrow x \in S$; Doesnt work $\Rightarrow$ Guess wrong.
Alice $S$-helpful $\Rightarrow$ Interpreter exists!

# Proof of Negative Result

- L not in PSPACE implies Bob makes mistakes.
  - Suppose Alice answers every question so as to minimize the conversation length.
    - (Reasonable effect of misunderstanding).
  - Conversation comes to end quickly.
  - Bob has to decide.
  - Conversation + Decision simulatable 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.

# Potential Criticisms of Main Theorem

- This is just rephrasing IP=PSPACE.
  - No ... the result proves “misunderstanding is equal to mistrust”. Was not a priori clear.
    - Even this is true only in some contexts.

# Potential Criticisms of Main Theorem

- This is just rephrasing IP=PSPACE.
- Bob is too slow: Takes exponential time in length of Alice, even in his own description of her!
  - A priori – not clear why he should have been able to decide right/wrong.
  - Polynomial time learning not possible in our model of “helpful Alice”.
  - Better definitions can be explored – future work.

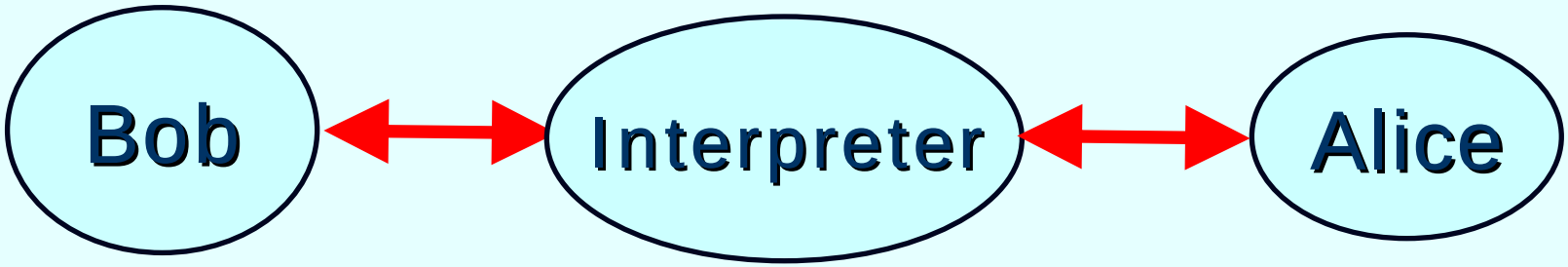
# Potential Criticisms of Main Theorem

- This is just rephrasing IP=PSPACE.
- Bob is too slow: Takes exponential time in length of Alice, even in his own description of her!
- Alice has to be infinitely/PSPACE powerful ...
  - But not as powerful as that Anti-Virus Program!
  - Wait for Part II

# Part II: Intellectual Curiosity

# Setting: Bob more powerful than Alice

- What should Bob's Goal be?
  - Can't use Alice to solve problems that are hard for him.
  - Can pose problems and see if she can solve them. E.g., Teacher-student interactions.
  - But how does he verify "non-triviality"?
  - What is "non-trivial"? Must distinguish ...




Scene 1

## Setting: Bob more powerful than Alice

- Concretely:
  - Bob capable of TIME($n^{10}$).
  - Alice capable of TIME($n^3$) or nothing.
  - Can Bob distinguish the two settings?

- Answer: Yes, if Translate(Alice,Bob) computable in TIME($n^2$).
  - Bob poses TIME($n^3$) time problems to Alice and enumerates all TIME($n^2$) interpreters.

- Moral: Language (translation) should be simpler than problems being discussed.

# Part III: Concluding thoughts

## Is this language learning?

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

# Goals of Communication

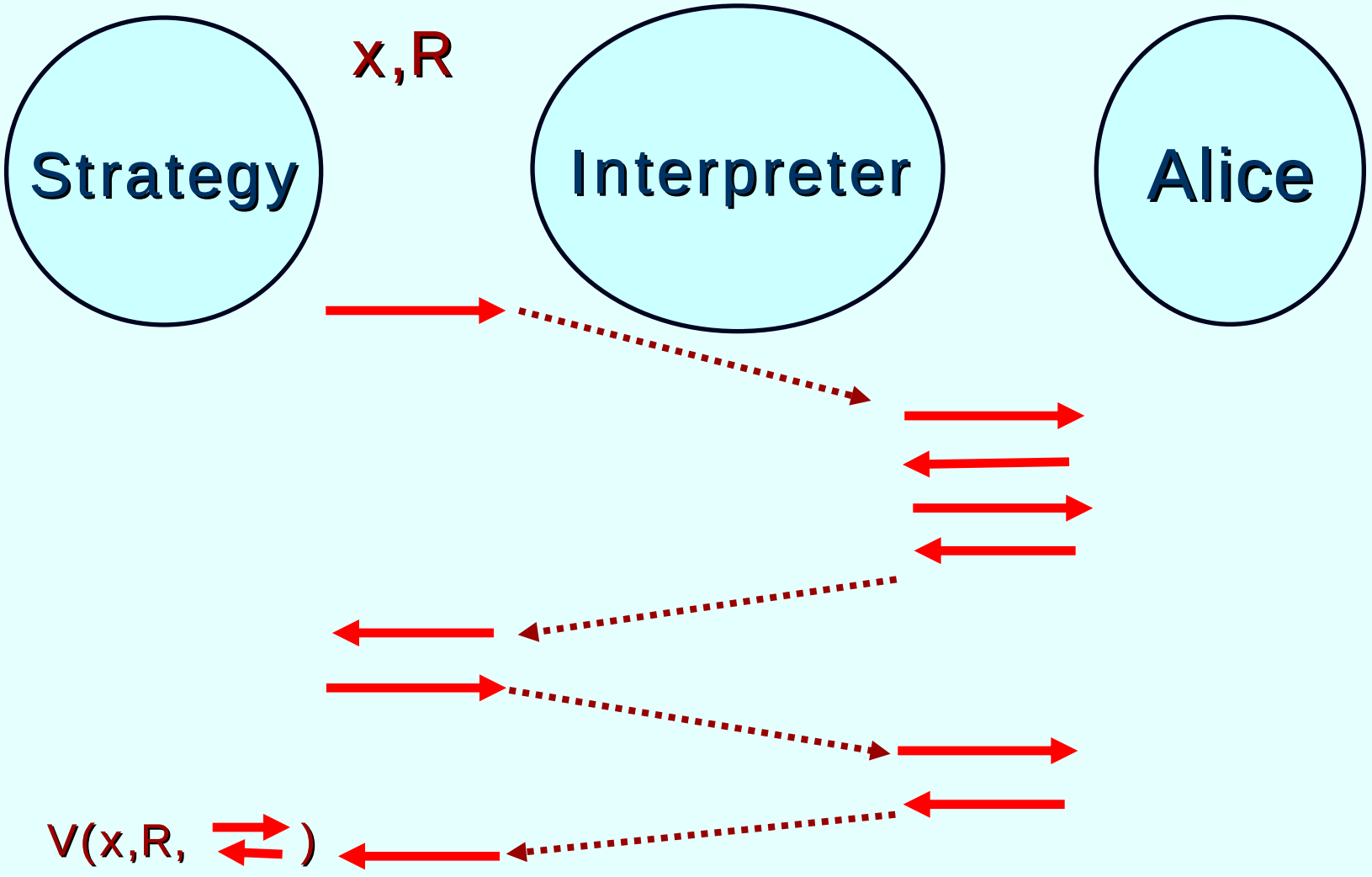
- Largely unexplored (at least explicitly)!
- Main categories
  - Remote Control:
    - Laptop wants to print on printer!
    - Buy something on Amazon
  - Intellectual Curiosity:
    - Learning/Teaching
    - Listening to music, watching movies
    - Coming to this talk
    - Searching for alien intelligence
  - May involve common environment/context.

# Extension to generic goals

- Generic (implementation of) Goal: Given by:
  - Strategy for Bob.
  - Class of Interpreters.
  - Boolean function G of
    - Private input, randomness
    - Interaction with Alice through Interpreter
    - Environment (Altered by actions of Alice)

- Should be
  - Verifiable: G should be easily computable.
  - Complete: Achievable w. common language (for some Alice, independent of history).
  - Non-trivial: Not achievable without Alice.

# Generic Verifiable Goal




Verifiable Goal = (Strategy, Class of Interpreters, V)

# Generic Goals

- **Can define** Goal-helpful; Goal-universal**; and prove** existence of Goal-universal Interpreter **for all Goals.**
- Claim: **Captures all communication**

  **(unless you plan to accept random strings).**
- **Modelling natural goals is still interesting. E.g.**
  - **Printer Problem:** Bob(x): Alice should say x.
  - **Intellectual Curiosity:** Bob: Send me a "theorem" I can't prove, and a "proof".
  - **Proof of Intelligence (computational power):**

    Bob: given f, x; compute f(x).
- Conclusion: **(Goals of) Communication can be achieved w/o common language**

# Role of common language?

- If common language is not needed (as we claim), then why do intelligent beings like it?
  - Our belief: To gain efficiency.
    - Reduce # bits of communication
    - # rounds of communication
- Topic for further study:
  - What efficiency measure does language optimize?
  - Is this difference asymptotically significant?

## Further work

- Exponential time learning (enumerating Interpreters)
  - What is a reasonable restriction on languages?
- What are other goals of communication?
- What are assumptions needed to make language learning efficient?

**Paper (Part I) available from ECCC**

Thank You!

# Example

- Symmetric Alice and Bob (computationally):
  - Bob's Goal:
    - Get an Interpreter in TIME($n^2$), to solve TIME($n^3$) problems by talking to Alice?
    - Verifiable: Bob can generate such problems, with solutions in TIME($n^3$).
    - Complete: Alice can solve this problem.
    - Non-trivial: Interpreter can not solve problem on its own.

# 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.

## 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 $x \in L$!

# 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?

# 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.

# 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: Pioneer Faceplate?
## Non-interactive proofs of intelligence?

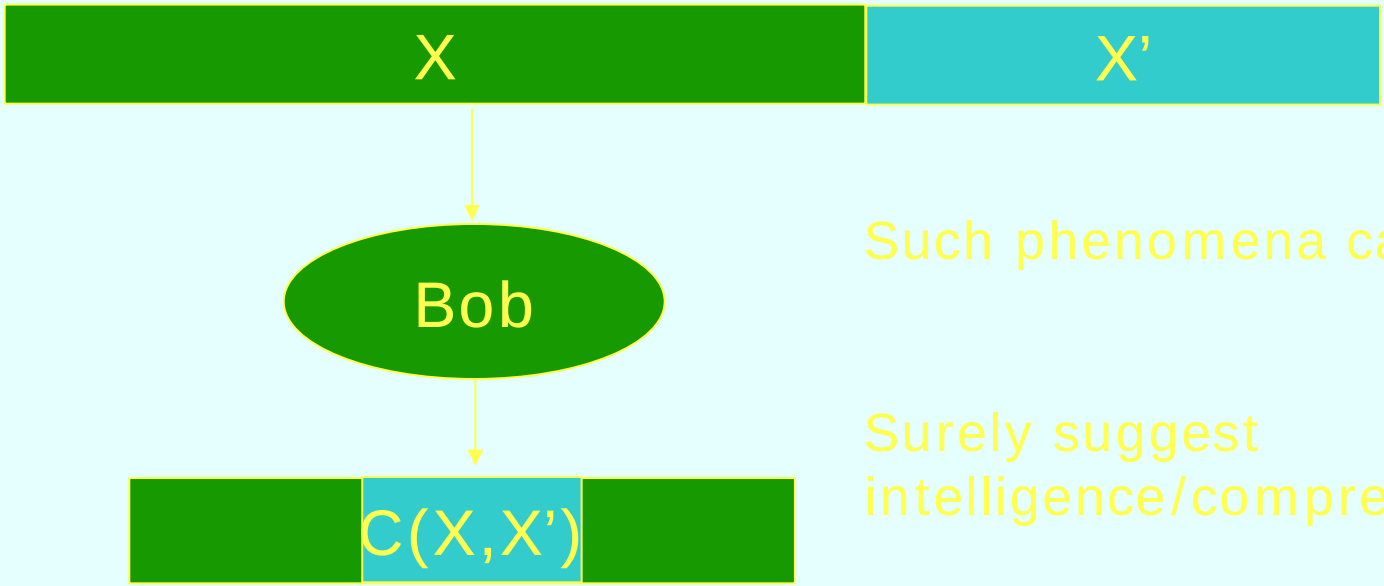
# Compression is universal

- 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 $D(C(X)) \neq X$ 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




Such phenomena can occur!

Surely suggest intelligence/comprehension?

## Discussion of result

- Alice needs to solve PSPACE. Realistic?
  - What about virus-detection? Spam-filters?
    - These solve undecidable problems!!!
  - PSPACE-setting: natural, clean setting
    - Arises from the proof.
    - Other languages work, (SZK, NP $\cap$ coNP).
- Learning B' is taking exponential time!
  - This is inevitable (minor theorem) ...
  - ... unless, languages have structure. Future work.

## Discussion of result

- Good news:
  - If we take “self-verification” as an axiom, then meaningful learning is possible!
  - Simple semantic principle ... reasonable to assume that Alice (the alien) would have determined this as well, and so will use this to communicate with us (Bobs).