

# Brief Announcement: A Theory of Goal-Oriented Communication\*

Oded Goldreich  
Weizmann Institute for  
Science  
Rehovot, Israel  
[oded.goldreich@weizmann.ac.il](mailto:oded.goldreich@weizmann.ac.il)

Brendan Juba<sup>†</sup>  
MIT CSAIL & Harvard SEAS  
Cambridge, MA, USA  
[bjuba@alum.mit.edu](mailto:bjuba@alum.mit.edu)

Madhu Sudan  
Microsoft Research  
Cambridge, MA, USA  
[madhu@mit.edu](mailto:madhu@mit.edu)

## ABSTRACT

We put forward a general theory of *goal-oriented communication*, where communication is not an end in itself, but rather a means to achieving some *goals* of the communicating parties. Focusing on goals provides a framework for addressing the problem of potential “misunderstanding” during communication, where the misunderstanding arises from lack of initial agreement on what protocol and/or language is being used in communication. Despite the enormous diversity among the goals of communication, we propose a simple model that captures *all* goals.

## Categories and Subject Descriptors

F.0 [Theory of Computation]: General; E.4 [Coding and Information Theory]: Formal models of communication

## General Terms

Theory

## 1. INTRODUCTION

Modern computer systems consist of pieces that are built at different times, by different groups, that make no special effort to coordinate with one another. Moreover, any “standards” employed by these parties change over time as the amount of data (and/or address spaces) grows and as new applications and features are introduced. Thus, there are no longer any guarantees that these pieces *interoperate*—and indeed, everyday experience shows that new components are often incompatible with the old. We therefore wish to understand when interoperability in the absence of fixed standards is possible and at what expense.

To our knowledge, though, all existing models of communication problems – building on the seminal work by Shannon [7] – implicitly assume that the communicating components are designed together, so the issues of (in)compatibility that we wish to study do not arise and cannot be captured in these models. This deficiency was pointed out in an earlier work by Juba and Sudan [3], who suggested that these

\*The full version of this work is available on ECCO [1] and in adapted form in the second author’s Ph.D. thesis [2].

<sup>†</sup>Supported by NSF Award CCF-0939370

problems might be addressed by explicitly introducing the *goal of the communication*, and furthermore demonstrated that for one example goal – delegating computation of a function – *universal interoperability* can be guaranteed. In this work, we establish that the earlier suggestion of Juba and Sudan is essentially valid by introducing a mathematical model of generic goal-oriented communication and showing that for broad classes of goals in this model, we can design components (algorithms) with universal compatibility.

In particular, the main feature of our work is the level of generality: while the earlier work of Juba and Sudan relied on a computational imbalance between the components, and in particular only considered the delegation of a PSPACE-complete function by a polynomial-time bounded algorithm, our work addresses any case in which there is something to gain from the communication. For example, the problem of using a printer to produce a document – which cannot be cast as a problem of delegating computation in any reasonable sense – is captured naturally by the simple model introduced in the current work.

## 2. THE MODEL OF GOAL-ORIENTED COMMUNICATION

We focus primarily on the case of communication between a pair of entities, since this is sufficient to capture the essential issues of incompatibility that may arise. Moreover, we focus on an asymmetric setting of communication between parties that we refer to as a *user* – who represents “us” or “our point of view,” and in any case, operates on our behalf – and a *server*, whose assistance towards achieving a *goal* we seek via communication.<sup>1</sup> In particular, we consider a synchronous model of communication in which the parties are described by *strategies* that take a *internal state* and an incoming message profile to a (distribution over) a new state and an outgoing message profile. Following the work of Juba and Sudan [3], the core of the problems of incompatibility that we consider are captured by considering, instead of a single server strategy, a *class* of possible server strategies: roughly, a user strategy is compatible with the entire class of servers if its goal of communication is achieved whenever the user is paired with any (adversarially selected) server in the class.

Note, however, that in our model as described thus far, neither the user strategy, nor the server strategy, and *cer-*

<sup>1</sup>We remark that the full version [1] briefly considers a symmetric setting with more than two parties, but this primarily consists of a reduction to the two-party setting.

tainly not the contents of the communications channels they share should be specified (fixed) by the goal of communication. The goal is thus introduced by introducing a *third* entity to the model, capturing either a hypothetical “referee,” “the rest of the system,” or “the environment,” which monitors the communication between the user and the server by communicating with them, that we refer to as the *world*. We postulate that the (user’s) goal for the communication can be described in terms of the states of this third party’s strategy. That is, to fix a goal of communication, we take the world’s (non-deterministic<sup>2</sup>) strategy as fixed, and fix a set of acceptable sequences of world states (or equivalently, define a *referee* predicate on the set of all possible histories of world states). We say that “the goal is achieved” if the system produces a sequence of world states that is acceptable.

At this level of generality, clearly not every goal can be achieved by some user-server pair, and our focus is naturally on cases in which the only issue is compatibility: that is, in which *some* reasonable (e.g., polynomial time) user strategy would achieve the goal with an adequate server. Actually, in order to simplify matters, we focus exclusively on *forgiving* goals in which every finite partial history can be extended to a successful history. Fixing such a forgiving goal, we say that a server strategy is *helpful* for the goal and a class of user strategies if there is some user strategy  $U$  such that when  $U$  is paired with the server, and the server and world are started from *any* initial state, the goal is achieved (i.e., the referee is satisfied). We then refer to a user strategy as *universal* for a goal (and a class of user strategies) if it achieves the goal with every server that is helpful for the goal (with respect to this class of user strategies).

### 3. NOTIONS AND RESULTS OF THE THEORY

As stated in the introduction, our second main contribution is that we are able to identify broad classes of goals of communication for which universal user strategies can be designed. These classes are described in terms of the feedback available to the user regarding its performance towards achieving the goal. Broadly speaking, we consider two families of goals, with different notions of feedback: one is *finite* goals, in which the user must halt at some point, and the referee is defined on these finite histories, and the other is *compact* goals in which the system runs for infinite time, and the referee’s decision is determined by whether the number of “unacceptable” prefixes of the history is finite or infinite. In each of these two cases, we introduce notions of feedback that we call *sensing*, which formally are predicates of the history of the portion of the system visible to the user. In each case, we specify two properties, called *viability* and *safety*, that the Boolean indications produced by sensing should satisfy, in order to be useful as a source of feedback. Loosely speaking, for compact goals, *viability* means that the user only obtains positive indications whenever it is coupled with a server that leads to achieving the goal (i.e., producing an acceptable execution), whereas *safety* means that negative indications are obtained when the user strategy is coupled with a server that does not lead

<sup>2</sup>Precisely, we find it convenient to assume that the world makes a single non-deterministic choice of a standard probabilistic strategy.

to achieving the goal. For finite goals, *safety* roughly means that positive indications are only obtained on acceptable histories, and *viability* means that a positive indication would be obtained by *some* (unknown) user strategy with a given server strategy.

Our main result asserts that whenever feedback (captured by safe and viable sensing) is available for a goal and a class of servers, there exists a user strategy (i.e., a universal one) that achieves the goal whenever coupled with any server in the class.

**THEOREM 1 (MAIN RESULT, LOOSELY STATED).** *For any (compact or finite) goal and any class of server strategies for which there exists safe and viable sensing, there exists a universal user strategy.*

In particular, if sensing is safe and viable with all helpful server strategies, then the theorem guarantees that we can design a user strategy that is universal with respect to any helpful server; that is, this universal strategy achieves the goal when coupled with a server  $S$  *iff* there is *some* user strategy that achieves the goal when coupled with  $S$ .

Loosely speaking, in the compact case, Theorem 1 is proved by enumerating all relevant user strategies and switching from the current strategy to the next one when a negative indication is obtained from the sensing function; in the finite case, strategies are enumerated “in parallel” as in Levin’s approach in [6], and sensing is used to decide when to stop. We mention that, in general, the overhead introduced by the enumeration is essentially necessary; that is, there exist natural cases in which any universal strategy must incur such an overhead. However, in special cases of interest, better performance may be possible. Indeed, we view our results as the first steps in a new direction, which motivate the search for algorithms that are compatible with broad classes (that may not include all helpful servers). We note that some examples of such algorithms have been subsequently identified by Juba and Sudan [4] and Juba and Vempala [5] (also described in [2]).

### 4. ACKNOWLEDGEMENTS

We thank Boaz Patt-Shamir for encouraging this announcement, and an anonymous reviewer for insightful comments.

### 5. REFERENCES

- [1] O. Goldreich, B. Juba, and M. Sudan. A Theory of Goal-Oriented Communication. Technical Report TR09-075, ECCC, 2009.
- [2] B. Juba. Universal Semantic Communication. PhD thesis, MIT, September 2010.
- [3] B. Juba and M. Sudan. Universal Semantic Communication I. In *40th STOC*, pages 123–132, 2008.
- [4] B. Juba and M. Sudan. Efficient Semantic Communication via Compatible Beliefs. In *2nd ICS*, pages 22–31, 2011.
- [5] B. Juba and S. Vempala. Semantic Communication for Simple Goals is Equivalent to On-line Learning. Manuscript, 2011.
- [6] L. A. Levin. Universal Search Problems. *Probl. Inform. Transm.*, 9:265–266, 1973.
- [7] C. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.