# Preface

## Abstract

Is meaningful communication possible between two intelligent parties who share no common language or background? We propose that this problem can be rigorously addressed by explicitly focusing on the goals of the communication. We propose a theoretical framework in which we can address when and to what extent such semantic communication is possible.

Our starting point is a mathematical definition of a generic *goal for communication*, that is pursued by *agents* of bounded computational complexity. We then model a "lack of common language or background" by considering a *class* of potential partners for communication; in general, this formalism is rich enough to handle varying degrees of common language and backgrounds, but the complete lack of knowledge is modeled by simply considering the class of all partners with which *some* agent of similar power could achieve our goal. In this formalism, we will find that for many goals (but not *all*), communication without any common language or background *is* possible. We call the strategies for achieving goals without relying on such background *universal protocols*.

The main intermediate notions introduced by our theory are formal notions of feedback that we call *sensing*. We show that in many natural settings of interest, sensing captures the essence of whether or not reliable universal protocols can be constructed: we find that across settings, sensing is almost always sufficient, usually necessary, and generally a useful design principle for the construction of universal protocols. We support this last point by developing a number of examples of protocols for specific goals. Notably, we show that *universal delegation of computation* from a space-efficient client to a general-purpose server is possible, and we show how a variant of TCP can allow end-users on a packet network to *automatically* adapt to small changes in the packet format (e.g., changes in IP).

The latter example above alludes to our main motivation for considering such problems, which is to develop techniques for modeling and constructing computer systems that do not require that their components strictly adhere to protocols: said differently, we hope to be able to design components that

function properly with a sufficiently wide range of other components to permit a rich space of "backwards-compatible" designs for those components. We expect that in the long run, this paradigm will lead to *simpler* systems because "backwards compatibility" is no longer such a severe constraint, and we expect it to lead to more *robust* systems, partially because the components should be simpler, and partially because such components are inherently robust to deviations from any fixed protocol.

Unfortunately, we find that the techniques for communication under the *complete* absence of any common background suffer from overhead that is too severe for such practical purposes, so we consider two natural approaches for introducing some assumed common background between components while retaining some nontrivial amount of flexibility. The first approach supposes that the designer of a component has some "belief" about what protocols would be "natural" to use to interact with other components; we show that, given sensing and some sufficient "agreement" between the beliefs of the designers of two components, the components can be made universal with some relatively modest overhead. The second approach supposes that the protocols are taken from some restricted class of functions, and we will see that for *certain* classes of functions and *simple* goals, efficient universal protocols can again be constructed from sensing.

Actually, we show more: the special case of our model described in the second approach above corresponds *precisely* to the well-known model of *mistake-bounded on-line learning* first studied by Bārzdiņš and Freivalds, and later considered in more depth by Littlestone. This connection provides a reasonably complete picture of the conditions under which we can apply the second approach. Furthermore, it also seems that the first approach is closely related to the problem of *designing good user interfaces* in Human-Computer Interaction. We conclude by briefly sketching the connection, and suggest that further development of this connection may be a potentially fruitful direction for future work.

## About this book

This book contains a revised edition of my Ph.D. dissertation, completed under the supervision of Madhu Sudan, and submitted to the Department of Electrical Engineering and Computer Science at MIT on August 30, 2010. As is typical for such works, portions of its contents have appeared (or will appear) in papers published elsewhere. The list of these other publications to date, along with the changes to the text since the submitted version, appears in the Bibliographic notes section below.

## Organization

Although the Introduction is certainly a good place to start, a reader who wishes to "jump in" may start with Chapter 2 followed by Chapter 6, in which the technical core of the work is developed, for "finite" and "infinite" goals, respectively. (The latter is more general, but also more technically involved.) Chapters 2–5 only address finite goals, while Chapters 6–9 are primarily focused on infinite goals.

Variations on this main theory are considered in Chapters 5 and 7. Chapter 5 refines the notions of Chapter 2 to address other kinds of resource limitations, such as limitations on the available memory or number of random bits available. Chapter 7 considers some other relaxations of the basic models, such as allowing our algorithms to sometimes fail at finite goals, or granting users the ability to initiate small "exploratory" goals, which respectively may broaden the kinds of goals that can be achieved, or improve the efficiency of solutions.

Potential applications (i.e., example goals) are the focus of Chapters 3 and 9, but also appear along the way in Chapters 5 and 6, and Chapters 4 and 8 concern ways to improve the efficiency of the schemes presented here.

## Prerequisites

The main requirement for understanding the work presented here is some basic familiarity with the theory of algorithms and/or computational complexity at the undergraduate level, for example, as covered by Sipser's textbook [141]. Some knowledge of calculus or algebra at an undergraduate level is also essential for some of the proofs.

While it would not have been reasonable to include a review of the above background, I have included some appendices covering any material that is useful or essential that a reader with this background may not have encountered (or may have forgotten), specifically a review of the basic notions and results from probability in Appendix A, and an introduction to interactive proofs in Appendix B. Some material on game theory that may well be completely unfamiliar to such a reader, which is used on occasion (in Chapters 4 and 9), is presented in Appendix C.

## Bibliographic notes

The main differences between this revised edition and the original version submitted to MIT (beyond the inclusion of the aforementioned Appendices) occur primarily in Chapters 5 and 9. Specifically, in Section 5.5.2, the original version achieved universal delegation of computation by logspace devices using a rather involved construction of public-coin logspace interactive proof systems for P-complete problems due to Goldwasser, Kalai, and Rothblum [73], together with a generic result (based on the work of Condon and Ladner [49]) that showed that any such proof system for P-complete problems

would be necessarily logspace-competitive. This has all been replaced by a relatively simple, direct construction of a (private-coin) logspace-competitive proof system (Theorem 5.27). On the other hand, the bounds claimed in Chapter 9 were simplified by the use of a cleaner analysis (as presented by Cesa-Bianchi and Lugosi [43]) of the algorithm Exp3.P.1 for the nonstochastic bandit problem than that originally presented by Auer et al. [10]. A complete presentation of this improved analysis appears in Appendix C. A few additional remarks on the schemes presented in Chapter 9 – specifically, viewing the schemes from a protocol-level (i.e., specification level) perspective, pointing out connections to IPsec, and remarking on the scope of their applicability – are also included in this version, in response to feedback I've received.

The foundations of the work presented here are based on one previously published work with Madhu Sudan [83], and a couple of technical reports with Madhu Sudan [84] and Oded Goldreich and Madhu Sudan [70]. More specifically, the introductory chapter closely follows the paper with Madhu Sudan [83], while the technical report with Madhu Sudan [84] contains an *early* version of Chapters 3 and 5 (and thus implicitly also Chapter 2). Chapter 6, on the other hand, is a modified version of the technical report with Oded Goldreich and Madhu Sudan [70], and the second half of Chapter 7 is also adapted from the same report. (Likewise, the initial section of Chapter 8 appeared in the aforementioned technical report.) More generally, the present form of Chapter 2 (in contrast to the version described in the technical report with Madhu Sudan [84]) and thus also the present formulation of this work benefitted immensely from the input from Oded Goldreich.

Following submission of the thesis, some of the work presented here appeared in conference papers: specifically there is a paper with Madhu Sudan [85] based on the main contents of Chapter 4, and a paper describing the work with Santosh Vempala [86] comprising the core of Chapter 8, as well as Section 4.4 (in Chapter 4).

The rest of the work in this book was previously unpublished. I have tried to note throughout when the work was particularly indebted to the contributions of others. In general though, *unless otherwise noted, the work presented in this book is joint work with Madhu Sudan.*

## Acknowledgements

It seems fitting to start with my parents.

Over the years at MIT, I've slowly come to realize that a graduate student couldn't wish for better parents. You see, my parents both entered doctoral programs, so they knew what the allure was, and what challenges lay ahead; at the same time, moreover, my parents both *dropped out* of graduate school with master's degrees, so there was no sense of pressure or unrealistic expectations. Finally, they demonstrated that even if things didn't work out for

me at MIT, it would still be possible to lead a rich, fulfilling life. It took me a long time to realize how uncommon that was.

On the subject of support, I also want to thank my wife, Angelina. As I write these acknowledgements, she's sitting next to me, plugging away at the research for her own thesis. The writing process over the last year or so has been a grueling, all-consuming, long, hard slog, but she's endured it about as well as anyone could. Her companionship and the sense of solidarity with it has prevented the writing process from turning isolating and lonely. Moreover, she's acted as my test audience more times than could reasonably be expected of anyone, and some of the material here has benefitted substantially from her careful feedback.

I'm also deeply indebted to my officemates, especially those that were here during the early days of this work (and long "middle years" of graduate school) – Swastik Kopparty, Paul Valiant, Guy Rothblum, and Ben Rossman, and later Jing Chen – for the feedback, sanity checks, and numerous helpful conversations they provided while the work presented here was taking form. Similarly, I'd like to thank the rest of my groupmates over the years, Sergey Yekhanin, Victor Chen, Elena Grigorescu, Shubhangi Saraf, Tali Kaufman, and Jakob Nordström for conversations, feedback, and support of various kinds. Thanks to Jacob Scott for keeping the pressure on me to go to the gym regularly back in the early days. (Seriously.)

I want to thank Lenore Blum for suggesting that I should try a REU—prior to that, I hadn't realized that the kind of problem-solving that I liked was called "research." I also want to thank Steven Rudich for his course 15-251, "Great Theoretical Ideas in Computer Science" (and also for his course on computational complexity, as a result of which, I no longer have any fear of hard problems). I grew immeasurably as a result of my involvement in 15-251, both as a student and as a TA.

Moreover, I'd also like to thank Steven Rudich for several interesting conversations about "alien communication," i.e., the subject of this work. I had influential conversations about this work with a number of other people, including Eran Tromer, Leslie Valiant, Leslie Kaelbling, Adam Kalai, and Bob Berwick. Eran and Adam have been enthusiastic about the work, and provided numerous helpful insights – Adam specifically helped to inspire the model of Chapter 4, and Eran more generally – while the conversations with Professors Valiant and Kaelbling helped point the way towards the connection to on-line learning in Chapter 8. Bob Berwick directed me to the relevant background in Philosophy of Language. I had some *particularly* helpful conversations with Santosh Vempala, and several of the results here are joint work with Santosh, as indicated. Likewise, I want to thank David Sontag and Dan Roy for their continued enthusiastic interest in the work, and their suggestion of the connection to PAC-Bayes that inspired Chapter 4. Finally I want to thank Ryan Williams for several conversations, for his continued interest and support in this work, and (perhaps most of all) for his willingness to read some of our early drafts.

I'd like to thank my committee members, Mike Sipser and Silvio Micali. Mike had two particularly rare qualities: first, he took the time to read the write-ups that I sent to him, and second, he was almost always available. Silvio, on the other hand, has given me substantial, highly relevant advice over the years, on this work, and more generally. Also, although relative to him I'm a mere novice at the art of a good definition, I owe much of what I know to some of the early courses I had with him.

Now, I wish to acknowledge the influence that Oded Goldreich has had throughout this work. To put it briefly, anywhere the formalism is clear and the terminology suggestive, it is due to Oded, and anywhere the work seems hasty and awkward, it is the result of my own attempt to fill in.[1] Chapter 6 closely follows a technical report on which Oded was the principal author [70], and the present version of the framework in Chapter 2 (and by extension, Chapter 5, as well as the rest of the thesis!) was in turn deeply influenced by the choices made by Oded in reformulating and extending my earlier work with Madhu (from [84]) to the infinite execution setting. I want to thank Oded, first for his vote of confidence in our work, and subsequently for his untiring efforts to make it comprehensible. A theoretical framework such as the one proposed here would be no good to anyone if no one could make sense of it!

I'd also like to acknowledge the influence Manuel Blum has had, both on the direction of this project, and on me as a researcher. It's fair to say that the two years I worked with Manuel at Carnegie Mellon shaped my taste and identity as a researcher, and probably the main reason I was originally interested in the questions considered in this work was that I was looking for another way to make progress on some of the questions that Manuel had raised while I was working with him. I've had the pleasure of conversations with Manuel about this work several times since, and they've all proved immensely helpful. Manuel also continues to be perhaps *the* most amazingly supportive person I have ever met, and I can't thank him enough.

Finally, it's an honor to acknowledge the influence, direction, support, and contributions of my advisor, Madhu Sudan. After working with Manuel, I was looking for another significant but colorful, wide open problem to work on, and so when Madhu suggested working on "the problem of establishing a common language," during my first year, I was hooked. I'm extremely grateful for his willingness to work closely with me on these problems, as well as for his support, advice, and encouragement over the long years since. So long, Madhu! And, thanks for all the sushi.

Brendan Juba

*Cambridge, Massachusetts*
*August, 2010*

[1] The reader may confuse this for modesty, but a simple comparison of Chapter 2 with the original technical report [84] should clear up any lingering doubts.

The work presented here was supported by an Akamai Presidential Fellowship, an NSF Graduate Research Fellowship, and NSF grants CCR-0514915 and CCF-0726525; the preparation of this book was supported by NSF grant CCF-0939370. The publication of this Springer edition was made possible by the cooperation and efforts of Ronan Nugent of Springer.

http://www.springer.com/978-3-642-23296-1

Universal Semantic Communication
Juba, B.
2011, XX, 400 p., Hardcover
ISBN: 978-3-642-23296-1