Communication Amid Uncertainty

Madhu Sudan
Microsoft Research New England
One Memorial Drive
Cambridge, MA, USA 02142
Email: madhu@mit.edu

***Abstract*—The classical theory of communication, starting with the work of Shannon, has always that assumed the meaning of the messages being exchanged is known to the sender and receiver. This assumption set aside a tricky issue and allowed the theory to focus on the more pressing engineering problem of the time - namely communicating the bits efficiently and reliably.**

In the current times, we see increasing evidence that this question can no longer be set aside. On the one hand, communication of the bits have become very reliable, so reliability is no longer the pressing concern today. On the other hand, increasingly these bits are operated on by computers or mechanical devices. In such settings it becomes essential that the computers and machines know what the bits mean. In this article intended to accompany a talk to be given at the workshop, we describe some of our attempts to extract the notion of meaning, and the challenges this task poses.

Meaning is best understood by focussing on the phenomenon of "misunderstanding", i.e., when the receiver does not understand what the sender says. Misunderstanding, in turn, seems to emerge principally from "uncertainty": Senders and receivers are uncertain about what the other knows/believes. We illustrate the problem in a simple setting, before moving on to describing our attempts to tackle the general complex task.

Based on joint works with Brendan Juba (Harvard), Oded Goldreich (Weizmann), Adam Kalai (MSR New England), and Sanjeev Khanna (U. Penn.)

I. Introduction

In this brief article we describe a series of recent works [4], [2], [5], [3], joint with Juba, Goldreich, Kalai and Khanna, that attempt to explore some phenomena in communication that are prevalent in "self-designed" communication systems (e.g., human communication), as opposed to "centrally-designed" systems (most current engineered communication systems. This article is meant to serve as a gentle introduction to the works above, and is thus written in a more casual style. Readers seeking a more formal and careful treatment are encouraged to read the original articles for more precise and accurate descriptions. With that disclaimer, lets move on to the subject at hand.

In his seminal work, Shannon [6] asserts "The semantic aspects of communication are irrelevant to the engineering problem." With this master stroke he sets aside an issue which is otherwise confusing and highlights what must have been the daunting task of the time: How to convert unreliable channels into reliable ones, efficiently? Sixty plus years later, we have made significant progress on the question of reliability and efficiency, so much so that it is quite reasonable to think of each channel as being capable of implementing a reliable channel. Since the original problem is no longer so daunting, this starts to highlight some of the "lesser" problems of the time, namely, the semantics!

So what is semantics? This could be the subject of a deep debate (and has been among philosophers and communication scholars), but for our purpose we could go back to the sentence succeeding the above quote of Shannon where he continues to say: "The significant aspect is that the actual message is one selected from a set of possible messages." Functionally this is the aspect that is brought into question by the "semantics" aspect. What if the sender and receiver are not in agreement on the set of messages? What if they prefer different encodings (into bits) of the different messages? Indeed what are messages, as opposed to the bits that encode them?

The classical theory of information blurs the distinction between messages and their representation by a sequence of bits, by allowing the encoder and decoder of a system to be jointly designed. While this is a reasonable expectation in any centrally designed system, many natural forms of communication are not so-centrally designed, and ever some modern engineered systems are not so. Is it possible to have a theory of information where such encoders and decoders are *not* jointly designed? Looking at "natural" systems around us, we see plenty of examples where indeed they are not jointly designed, or at least not completely so. Human communication is a perfect example: A child is instilled at birth with some basic facilities that enable it to communicate, but it develops a more and more sophisticated codebook (namely the language it learns and the vocabulary it acquires). There are other examples, communication between humans and animals where it could be argued the the understanding is not perfect, but at the same time manages to convey lots of information. So one motivation for studying this "semantics" of information is to understand such communication systems, within a mathematical umbrella.

A second motivation nowadays is the ever increasing merger of communication and computation devices. Increasingly these merged devices want to interoperate freely — so we see more and more settings where the the two endpoints were not "jointly designed" (this is certainly true about the hardware, it becomes increasingly true about even the software interfaces). Such systems tend to work in practice today, but without a theory it is impossible to say anything about how they will work tomorrow.

With these motivations in mind, we now turn to some of the models, problems, and (our attempts at) solutions.

II. Modelling Uncertainty

In order to introduce "uncertainty" (of sender about receiver and vice versa), we need to introduce a simple change to the standard (Shannon) model of communication. Recall that in the standard model, we have a sender S sending a message $m \in M$ to a receiver R over a (possibly noisy) channel C. All the concepts in the previous sentence that were represented by capital letters are assumed to be public knowledge. So both sender and receiver known S, M, R and C. To change this model we now allow senders and receivers to themselves be element of a set.

a) Basic Model of Uncertainty:: In this setting a sender $s \in S$ wishes to send a message $m \in M$ to a receiver $r \in R$ over a channel C. While both sender and receiver are aware of S, R, M and C, they are no longer certain about each other. Only the sender s knows m and more importantly its own identity s, while only the receiver knows $r \in R$.

Thus in particular s does not know r and vice versa, and has to overcome this diversity in its audience. We assume an adversary A picks the pair (s, r) and asks them to communicate with each other, and the goal could be to ensure that the receiver decodes m correctly, while using the channel efficiently. (We will discuss generalizations of goals in later sections, but this can be the starting point.)

What distinguishes different $s \in S$, or different $r \in R$? It could be their behavior - different $s \in S$ may behave differently (encode information differently etc.). In such case there is not much of a "design" question left - every element has already been prescribed and the research question is usually to analyze the performance of a given setting, and possibly to identify the "good" pairs $(s, r) \in S \times R$ which work well, or more ambitiously to find $s \in S$ that works well for all $r \in R$ are vice versa. (Working well is defined to be achievement of the goal, which for now can be taken to imply that receiver decodes m correctly, and the channel is used efficiently.) Once again as implied by our notation we assume in such a case that the adversary A is known to the sender and the receiver, so if A never chooses some pair (s, r) it is not required that the system works well for such a pair (s, r).

A different reason for difference in $s \in S$ is their priors, or their past knowledge/experience, but it does not specify their behavior. In such a case their behavior, given their prior, may be left to the designer of the system, and then the research question will be to design encoders $E(\cdot, \cdot)$ and decoders $D(\cdot, \cdot)$ such that if the sender acts according to $E(s, \cdot)$ and receiver acts according to $D(r, \cdot)$ then the system works well. (In our example goal, if the sender sends $x = E(s, m)$, the channel outputs $y = C(x)$ and receiver decodes $\hat{m} = D(s, y)$ we would want $m = \hat{m}$ with high probability.)

In following sections we will start with the model above to pose some questions about some themes in communication which don't fit the standard model of communication, and illustrate some of the challenges.

III. Example: Compression

We start with the most simple problem in this setting, namely "source coding" aka compression. (The problem and results of this section are from [3].) In the usual definition of the problem, as say, solved by Huffman coding, the sender and receiver know some distribution P on the message space M and have to encode the message so that the expected length of the encoding of the message, when the message is drawn from P, is as small as possible. The classical solution leads to an encoding of length at most $H(P) + O(1)$, where $H(P)$ is the entropy of the distribution P.

Uncertainty between sender and receiver models a simple phenomenon. Sender and receiver are not in agreement over what distribution the message comes from. So in our setting the sender believes the distribution over messages is some P, while the receiver believes the distribution is some Q. Even if P and Q are very close, but the sender and receiver are not knowledgeable about each other, standard solutions fail when encoding/decoding. Below we will describe this resulting problem mathematically in a self-contained way.

We use $[n]$ to denote the set $\{1, \ldots, n\}$ and $\{0, 1\}^* = \cup_{n \geq 0} \{0, 1\}^n$. We use $\mathbf{Exp}_{X \leftarrow D}[X]$ to denote the expectation of a random variable drawn from some distribution D. For $x \in \{0, 1\}^*$, we use $|x|$ to denote its length.

Problem 3.1: Let $\Omega(M)$ denote the space of all possible probability distributions on M. (Note $\Omega(M)$ can be represented as a subset of $\Re^{|M|}$.) Let $A \subseteq \Omega(M) \times \Omega(M)$. Design an encoder $E : \Omega(M) \times M \to \{0, 1\}^*$ and decoder $D : \Omega(M) \times \{0, 1\}^* \to M$ such that the following hold:

Correctness: For every $(P, Q) \in A$ and every $m \in M$ $D(Q, E(P, m)) = m$.

Efficiency: For every $(P, Q) \in A$, $\mathbf{Exp}_{m \leftarrow P}[|E(P, m)|] \leq L_A(P)$.

The above problem really represents a family of problems, one for each adversary A. The solutions are thus characterized by the expected compression length $L_A(\cdot)$ that they achieve. To help digest these notions, let us consider the classical communication problem (where sender and receiver agree on P). This setting is modelled by the adversary $A = \{(P, P) | P \in \Omega(M)\}$, and the solution achieved here is given by $L_A(P) = H(P) + 1$.

Our challenge now is to design E and decoding schemes for other possible adversaries, in particular those that are not just diagonal. The work [3] considers a broad class of adversaries that we define next. To do so, we need a (new) notion of distance between distributions: Define $\Delta_{\max}(P, Q) = \max\{\Delta^L_{\max}(P, Q), \Delta^R_{\max}(P, Q)\}$ where $\Delta^L_{\max}(P, Q) = \max_{m \in M}\{\log_2 \frac{P(m)}{Q(m)}\}$ and $\Delta^R_{\max}(P, Q) = \max_{m \in M}\{\log_2 \frac{Q(m)}{P(m)}\}$. (Assume $0/0 = 1$ for the definition above.) Note that $\Delta_{\max}$ is a metric. It is really a symmetrized, worst-case divergence measure between P and Q. In particular, if Δ^L had been defined as the expectation, rather than the max, of $\log_2 P/Q$, then it would have been the divergence between P and Q, (and Δ^R would have reversed the role of

P and Q). By making it symmetric and taking the max, we get a metric.

Returning to the problem at hand, now consider the adversary $A_\Delta = \{(P, Q) | \Delta_{\max}(P, Q) \leq \Delta\}$. This adversary picks P and Q with some variation being allowed, but bounds the variation, by bounding the distance. [3] show, that there is a randomized encoding and decoding scheme (with "shared randomness") that manages to achieve $L_{A_\Delta}(P) \leq H(P)+2\Delta+C$ for some universal constant C. In particular, if $\Delta = 0$ one recovers the classical problem and the solution is nearly optimal (off by at most C). But now the solution degrades gracefully even as Δ increases. We won't clarify what "shared randomness" means but it does weaken the nature of the answer, so a deterministic solution would be nicer.

IV. Towards Semantics: Goal-oriented communication

The problem of uncertainty, of course, goes beyond, uncertainty of communicating entities about each other's knowledge/beliefs. The more challenging setting is when players cannot assume any common design element (such as the encoder and decoder of the previous section). This is the setting that is most reflective of "misunderstanding", and in this section and the next we attempt to formalize the problems here and the solutions.

A concrete example might illustrate the issues we are alluding to better. So lets say you are in a new city where the local residents speak a language unknown to you. You wish to reach a particular destination. You could talk to one of the residents to get directions, but while the interaction may easily result in many bits being exchanged, your goal is really to convey your destination to the local, and to understand the directions they provide.

Notice that part of the issue is the uncertainty of the two communicating entities about each other. Depending on what language the locals speak, they give directions differently. Similarly depending on what language you, as the tourist, speak you may pose your queries differently. For each communicating entity there is some underlying map from words to their meaning and it is this map that is unknown to the other, and this is an "uncertainty" phenomenon. (Each map from word to meaning defines a different tourist, and each such map also defines a different locality.)

Given the nature of the uncertainty above, it may make sense to ask, is it possible for the two agents to "learn" each other's language, i.e., the mapping from words to their meaning? In this completely general sense the task turns out to be impossible. Certainly there are many words in any language that would not appear in any conversation where one is asking for directions. And such words, since they were never used would never be learned. On the other hand, failure to learn the meaning of such words is not a communication failure! After all such words are not needed for the specific goal on this interaction, which was for you to learn how to get to a specific destination.

The general approach in [4], [2] is based on the following thesis.

Thesis: *Communication is not an end in itself. Rather it is a means to achieving some end goal.*

Focussing on the goal, gives a functional basis for "meaning". Communication in a goal-oriented setting is defined to be semantically successful if it helps the communicating entities achieve their goal. Your goal above was to get directions, and the local resident's goal might have been to just help you. With such compatible goals it is a reasonable expectation that communication can lead to satisfaction for both parties, and this ought to be the basis of semantic communication.

Identifying and focussing on the goals turns out to be a very useful task not only in defining semantic communication, but also enables, with some restrictions, achievement of successful semantic communication. The restrictions roughly state that progress towards the goal should be something that the communicating parties should be able to *sense*. In our example, this would roughly correspond to the following: You should at least be able to know when you've reached your destination. (Even better would be if you could estimate the distance, at least crudely.) Clearly this is a necessary condition: If you can't even recognize your destination when you reach it, you could take any translation of the resident's words to directions and this would take you somewhere, and you would not know that something went wrong. Thus having a goal whose achievement you can detect seems like a necessary condition even for *detecting errors*. Correcting errors is trickier and no universally efficient solution is known — the only recourse known is that you should try all possible interpretations of the instructions provided till you find an interpretation that leads you to the destination.

Trying out all possible interpretation may seem inefficient, but you should recall that the instructions probably work with a very limited vocabulary, and a sufficiently intelligent local resident would try not to use complex directions such as "Walk straight for 375 metres, avoid all potholes, and then look for the tallest building around you and head towards it." More reasonably they would say "Straight, straight, straight, then right" (so the vocabulary is limited and with a preference to repeating words).

The above attempts to give a very informal description of the nature of the results in [4], [2], [5], which attempt to give mathematical formalisms and some quantitative measures to describe the semantic communication problem and solutions. Specifically the work [4] picks a very specific goal of communication (which happens to be purely informational, and does not depend on the environment, or the local geography etc.) and shows that (1) the goal can not be achieved without communication, (2) it can be achieved with communication, provided the players can sense progress towards the goal, and (3) it can not be achievd even with communication, in the presence of sufficient uncertainty, if the players cannot sense progress towards the goal. The work [2] starts with the example from [4] and builds a general theory of goal-oriented communication. Such a theory requires defining (a) general communicating entities, (b) uncertainty for such entities, (c) general goals which remain fixed even when the adversary changes the players, and (d) sensing progress. With these definitions the work manages to reproduce the results (1)-(3) described above in this general setting. Finally, [5] attempts to build some quantitative measures of compatibility between the two communicating players and show how the communication can be made more efficient as players become more compatible (somewhat akin to the results described in Section III).

V. Conclusion

The world of natural communication abounds with intriguing phenomena that sometimes resemble the phenomena seen in designed communication, and at other times seem to be at odds with them. An intriguing element in natural communication is "language" — a codebook that is produced by distributed design and evolutionary principles, rather than by an efficient designer. Yet, it does not preclude elements of efficiency, harmony, modularity. Periodically languages have been streamlined by the introduction of grammars, dictionaries, and by structured education. Among the many forces at work in the evolution of language are a mix of information-theoretic efficiency considerations (in particular source coding), computational considerations (languages should be easy to learn and remember) and the need for reliability against uncertainty, and/or diversity of the audience. Such phenomena are sure to happen also in the engineered world as we move towards increasing diversity of devices. Our hope is that even in this future world, principles of information theory, and theory of computing, and in general mathematically based analyses, will continue to play a central role in guiding such evolutions.

Acknowledgment

I would like to thank Brendan Juba, Oded Goldreich, Adam Kalai and Sanjeev Khanna for the collaborations. I would also like to thank Gireeja Ranade, Elad Haramaty, David Tse, and Sudeep Kamath for their comments on this line of work.

References

[1] Bernard Chazelle, editor. *Innovations in Computer Science - ICS 2010, Tsinghua University, Beijing, China, January 7-9, 2011. Proceedings*. Tsinghua University Press, 2011.

[2] Oded Goldreich, Brendan Juba, and Madhu Sudan. A theory of goal-oriented communication. *J. ACM*, 59(2):8, 2012.

[3] Brendan Juba, Adam Tauman Kalai, Sanjeev Khanna, and Madhu Sudan. Compression without a common prior: an information-theoretic justification for ambiguity in language. In Chazelle [1], pages 79–86.

[4] Brendan Juba and Madhu Sudan. Universal semantic communication I. In *Proceedings of the 2008 ACM International Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 123–132. ACM, 2008.

[5] Brendan Juba and Madhu Sudan. Efficient semantic communication via compatible beliefs. In Chazelle [1], pages 22–31.

[6] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.