Electronic Colloquium on Computational Complexity, Report No. 64 (2015)

Communication with Contextual Uncertainty

Ilan Komargodski * Pravesh Kothari † Madhu Sudan ‡

Abstract

We introduce a simple model illustrating the role of context in communication and the challenge posed by uncertainty of knowledge of context. We consider a variant of distributional communication complexity where Alice gets some information x and Bob gets y, where (x, y) is drawn from a known distribution, and Bob wishes to compute some function $g(x, y)$ (with high probability over (x, y)). In our variant Alice does not know g, but only knows some function f which is an approximation of g. Thus, the function being computed forms the context for the communication, and knowing it imperfectly models (mild) uncertainty in this context.

A naive solution would be for Alice and Bob to first agree on some common function h that is close to both f and g and then use a protocol for h to compute $h(x, y)$. We show that any such agreement leads to a large overhead in communication ruling out such a universal solution. In contrast, we show that if g has a one-way communication protocol with low complexity in the standard setting, then it has a low communication protocol (with a constant factor blowup in communication and error) in the uncertain setting as well. We pose the possibility of a two-way version of this theorem as an open question.

*Weizmann Institute of Science, Israel. Email: ilan.komargodski@weizmann.ac.il. Work done while an intern at MSR New England. Supported in part by a grant from the I-CORE Program of the Planning and Budgeting Committee, the Israel Science Foundation, BSF and the Israeli Ministry of Science and Technology.

†UT Austin, USA. Email: kothari@cs.utexas.edu. Work done while an intern at MSR New England.

‡Microsoft Research, One Memorial Drive, Cambridge, MA 02142, USA. Email: madhu@mit.edu.

ISSN 1433-8092

1 Introduction

Most forms of communication involve communicating players that share a large common context and use this context to compress communication. In natural settings the context may include understanding of language, and knowledge of the environment and laws. In designed (computer-to-computer) settings this includes knowledge of the operating system, communication protocols, and coding/decoding mechanisms. Remarkably, especially in the natural setting, context can seemingly be used to compress communication, even when it is not shared perfectly. This ability to communicate despite a major source of uncertainty has led to a series of works attempting to model various forms of communication amid uncertainty, starting with Goldreich, Juba and Sudan [JS08, GJS12] followed by [JKKS11, JS11, JW13, HS14, CGMS14]. This current work introduces a new theme to this series of works by introducing a functional notion of uncertainty and studying this model. We describe our model below and then contrast it with some of the previous works.

Model. Our model builds upon the classical model of communication complexity due to Yao [Yao79] and we develop it slowly here. The classical model considers two interacting players Alice and Bob each possessing some private information x and y with x known only to Alice and y to Bob. They would like to compute some joint function $g(x, y)$ and would like to do so while exchanging the minimum possible number of bits. In this work we suggest that the function g is the *context* to the communication and consider a setting where it is shared imperfectly. Specifically, we say that Bob knows the function g and Alice knows some approximation f to g. We ask when can Alice and Bob interact to compute $g(x, y)$ with limited communication?

It is clear that if $x \in \{0,1\}^n$, then n bits of communication suffice — Alice can simply ignore f and send x to Bob. We wish to consider settings that improve on this. To do so correctly on every input, a necessary condition is that g must have low-communication complexity in the standard model. However, this necessary condition does not appear sufficient — since Alice only has an approximation f to g. Thus we settle for a weaker goal: We settle for determining g correctly only on most inputs. This puts us in a distributional communication complexity setting. A necessary condition now is that g must have a low-error low-communication protocol in the standard setting. We now ask if g can be computed with low error, with low communication complexity, when Alice only knows an approximation f to g.

More precisely, in this setting the input to Alice is a pair (f, x) and to Bob is a pair (g, y). The functions (f, g) are adversarially chosen subject to the restrictions that they are close to each other (under some distribution μ on the inputs) and that g (and hence f) has a low-error low-communication protocol. The pair (x, y) is drawn from the distribution μ (independent of the choice of f and g). The players both know μ in addition to their respective inputs. (In some of the protocols below we will allow them to use shared randomness for conceptual simplicity, though in the distributional setting this is not necessary to attain a desired communication complexity.)

Results. To describe our results, we use a little notation. Let $\delta_\mu(f, g)$ denote the (weighted and normalized) Hamming distance between f and g with respect to the distribution μ. Let $\mathsf{CC}^\mu_\epsilon(f)$ denote the minimum communication complexity of a protocol that computes f correctly on all but ϵ fraction of the inputs. Let $\mathsf{owCC}^\mu_\epsilon(f)$ denote the corresponding *one-way* communication complexity of f. Given a family $\mathcal{F}$ of pairs of functions (f, g), we denote the uncertain complexity $\mathsf{CCU}^\mu_\epsilon(\mathcal{F})$ to be the minimum over all protocols Π of the maximum over $(f, g) \in \mathcal{F}$ and (x, y) in the support

of μ of the communication complexity of $\Pi((f,x),(g,y))$, subject to the condition that for every $(f,g) \in \mathcal{F}$, Π outputs $g(x,y)$ with probability $1-\epsilon$ over the choice of (x,y). That is,

$$\mathsf{CCU}^{\mu}_{\epsilon}(\mathcal{F}) \triangleq \min_{\{\Pi \mid \forall (f,g)\in\mathcal{F}:\, \delta_{\mu}(\Pi,g)\leq\epsilon\}} \max_{\{(f,g)\in\mathcal{F},(x,y)\in\mathsf{supp}(\mu)\}} \{\text{Comm. complexity of } \Pi((f,x),(g,y))\}.$$

Similarly, let $\mathsf{owCCU}^{\mu}_{\epsilon}(\mathcal{F})$ denote the one-way uncertain communication complexity of $\mathcal{F}$.

Our main theorem (see Theorem 4.1) shows that if μ is a product distribution on which f, g have low one-way communication complexity and f and g are close, then the pair (f,g) also has low (one-way) uncertain communication complexity. More precisely, let $\mathsf{ow}\mathcal{F}_{k,\epsilon,\delta}$ denote the family of all pairs of functions (f,g) with $\mathsf{owCC}^{\mu}_{\epsilon}(f), \mathsf{owCC}^{\mu}_{\epsilon}(g) \leq k$ and $\delta_{\mu}(f,g) \leq \delta$. Then, for every $\theta > 0$ we have $\mathsf{owCCU}^{\mu}_{\epsilon+2\delta+\theta}(\mathsf{ow}\mathcal{F}_{k,\epsilon,\delta}) = O(k/\theta^2)$.

Our result is significant in that it achieves (moderately) reliable communication despite uncertainty about the context, even when the uncertainty itself is hard to resolve. To elaborate on this claim, note that one hope for achieving a low-communication protocol for g would be for Alice and Bob to first agree on some function q that is close to f and g, and then apply some low-communication protocol for this common function q. This would be the "resolve the uncertainty first" approach. We prove (see Theorem 3.2) that resolving uncertainty can be very expensive (much more so than even the trivial protocol of sending x) and so this would not be a way to achieve Theorem 4.1. Instead we show a path around the inherent uncertainty to computing the desired function and this leads to a proof of Theorem 4.1.

Contrast with prior work. The first works to consider communication with uncertainty in a manner similar to this work were those of [JS08, GJS12]. Their goal was to model an extreme form of uncertainty, where Alice and Bob do not have any prior (known) commonality in context and indeed both come with their own "protocol" which tells them how to communicate. So communication is needed even to resolve this uncertainty. While their setting is thus very broad, the solutions they propose are much slower and typically involve resolving the uncertainty as a first step.

The later works [JKKS11, HS14, CGMS14] tried to restrict the forms of uncertainty to see when it could lead to more efficient communication solutions. For instance, Juba et al. [JKKS11] consider the compression problem when Alice and Bob do not completely agree on the prior. This introduces some uncertainty in the beliefs and provides fairly efficient solutions by restricting the uncertainty to a manageable form. Canonne et al. [CGMS14] were the first to connect this stream of work to communication complexity, which seems to be the right umbrella to study the broader communication problems. The imperfectness they study is however restricted to the randomness shared by the communicating parties, and does not incorporate any other elements. They suggest studying imperfect understanding of the function being computed as a general direction though they do not suggest specific definitions, which we do here.

Discussions and future directions. Arguably the model considered in this work is a fairly realistic one: Communication has some goals in mind which we model by letting Bob be interested in a specific function of the joint information that Alice and Bob possess. Moreover, it is a fairly natural model to posit that the two are not in perfect synchronization about the function that Bob is interested in, but Alice can estimate the function in some sense. The main weakness in the modeling is the specific notion of distance we use in our theorems that describes the gap between Bob's function and Alice's estimate. Here we made a simple choice via Hamming distance, which

forms a good first starting point. We believe it is interesting to propose and study other models of distance between functions that capture natural forms of uncertainty.

In terms of results, the fact that any one-way communication protocol can be carried out in the presence of uncertainty with only constant factor losses is a refreshingly positive statement. The two main restrictions here are the "one-way-ness" and the fact that Alice and Bob's input have to be from a product distribution. Neither restriction seems inherent — or at least, we do not see any reasons to believe so. We propose both as open questions: It would be nice to extend Theorem 4.1 to the two-way setting, or at least a bounded-round two-way setting. Our positive result works by establishing a canonical form for any one-way protocol in the standard distributional communication model: To compute $g(x, y)$, Alice tries to describe the entire function $g(x, \cdot)$ to Bob, and this does not create a huge overhead in communication. We are unable to show any canonical form for two-way communication setting and indeed it may not even exist. Deciding whether canonical forms that somehow change gradually as we morph from g to f seems to be the essence of the challenge in extending our results to the two-way setting. Similarly it would be very interesting to extend the setting to non-product distributions: we feel this would capture many natural settings compared to simple uniform or product distributions.

Finally, we wish to emphasize the mix of adversarial and probabilistic elements in our uncertainty model — the adversary picks (f, g) whereas the inputs (x, y) are picked from a distribution. We believe richer mixtures of adversarial and probabilistic elements could lead to broader settings of modeling and coping with uncertainty — the probabilistic elements offer efficient possibilities that are often immediately ruled out by adversarial choices; whereas the adversarial elements prevent the probabilistic assumptions from being too precise.

2 The Model

We start by recalling the classical communication complexity model of Yao [Yao79] and then present our definition and measures.

2.1 Communication Complexity

We start with some basic notation. For an integer $n \in \mathbb{N}$ we denote by $[n]$ the set $\{1, \ldots, n\}$. We use $\log x$ to denote a logarithm in base 2. For a distribution μ we denote by $x \sim \mu$ the process of sampling a value x from the distribution μ. Similarly, for a set X we denote by $x \sim X$ the process of sampling a value x from the uniform distribution over X. For any event E, let $\mathbb{1}(E)$ be the 0-1 indicator of E. For a probability distribution μ over $X \times Y$ we denote by μ_X the marginal of μ over X. By $\mu_{Y|x}$ we denote the conditional distribution μ over Y conditioned on $X = x$.

Given a distribution μ supported on X and functions $f, g : X \to \Sigma$, we let $\delta_\mu(f, g)$ denote the (weighted and normalized) Hamming distance between f and g, i.e., $\delta_\mu(f, g) \triangleq \Pr_{x\sim\mu}[f(x) \neq g(x)]$. (Note that this definition extends naturally to probabilitistic functions f and g – by letting $f(x)$ and $g(x)$ be sampled independently.)

We now turn to the definition of *communication complexity*. A more thorough introduction can be found in [KN97]. Let $f\colon X \times Y \to \{0, 1\}$ be a function and Alice and Bob be two parties. A protocol Π between Alice and Bob specifies how and what Alice and Bob communicate given their respective inputs and communication thus far. It also specifies when they stop and produce an output (that we require to be produced by Bob). A protocol is said to be one-way if it involves a

single message from Alice to Bob, followed by Bob producing the output. Π is said to compute f if for every $(x, y) \in X \times Y$ it holds that $\Pi(x, y) = f(x, y)$. The communication complexity of Π is the number of bits transmitted during the protocol between Alice and Bob. The communication complexity of f is the minimal communication complexity needed for a protocol to compute f.

It is standard to relax the above setting by introducing a distribution μ over the input space $X \times Y$ and requiring the protocol to succeed with high probability (rather than with probability 1). It is also standard to provide Alice and Bob a shared random string which is independent of x, y and f, but such randomness is not needed to minimize the distributional communication complexity. We say that a protocol Π ϵ-computes a function f under distribution μ if $\delta_\mu(\Pi(x, y), f(x, y)) \leq \epsilon$.

Definition 2.1 (Distributional Communication Complexity)**.** *Let $f\colon X \times Y \to \{0, 1\}$ be a Boolean function and μ be a probability distribution over $X \times Y$. The* distributional communication complexity *of f under μ with error ϵ, denoted by $\mathsf{CC}^\mu_\epsilon(f)$, is defined as the minimum over all protocols Π that ϵ-compute f over μ, of the communication complexity of Π. The one-way communication complexity $\mathsf{owCC}^\mu_\epsilon(f)$ is defined similarly minimizing over one-way protocols Π.*

In this paper, unless stated otherwise, whenever we refer to a protocol, we think of it being in the distributional communication model.

2.2 Uncertain Communication Complexity

We now turn to the central definition of this paper, namely *uncertain communication complexity*. Our goal is to understand how Alice and Bob can communicate when the function that Bob wishes to determine is not known to Alice. In this setting, we make the functions g (the function that Bob wants to compute) and f (Alice's estimate of g) explicitly part of the input to the protocol Π. Thus, in this setting a protocol Π specifies how Alice with input (f, x) and Bob with input (g, y) communicate, and how they stop and produce an output. We say that Π computes (f, g) if for every $(x, y) \in X \times Y$, the protocol outputs $g(x, y)$. We say that Π ϵ-computes (f, g) over μ if $\delta_\mu(g, \Pi) \leq \epsilon$.

Next, one may be tempted to define the communication complexity of a pair of functions (f, g) as the minimum over all protocols that compute (f, g) of their maximum communication. But this does not capture the uncertainty! (Rather a protocol that works for the pair corresponds to both Alice and Bob knowing both f and g.) To model uncertainty we have to consider the communication complexity of a whole class of pairs of functions, from which the pair (f, g) is chosen (in our case by an adversary).

Let $\mathcal{F} \subseteq \{f\colon X \times Y \to \{0, 1\}\}^2$ be a family of pairs of Boolean functions with domain $X \times Y$. We say that a protocol Π ϵ-computes $\mathcal{F}$ over μ if for every $(f, g) \in \mathcal{F}$, we have that Π ϵ-computes (f, g) over μ. We are now ready to present our main definition.

Definition 2.2 (Contextually Uncertain Communication Complexity)**.** *Let μ be a distribution on $X \times Y$ and $\mathcal{F} \subseteq \{f\colon X \times Y \to \{0, 1\}\}^2$. The communication complexity of $\mathcal{F}$ under contextual uncertainty, denoted $\mathsf{CCU}^\mu_\epsilon(\mathcal{F})$, is the minimum over all protocols Π that ϵ-compute $\mathcal{F}$ over μ, of the maximum communication complexity of Π over $(f, g) \in \mathcal{F}$ and (x, y) from the support of μ.*

As usual the one-way contextually uncertain communication complexity $\mathsf{owCCU}^\mu_\epsilon(\mathcal{F})$ is defined similarly.

Observe that in the special case where $\mathcal{F} = \{(f, g)\}$, Definition 2.2 boils down to the standard definition of distributional communication complexity (see Definition 2.1) for function g,

and so we have $\mathsf{CCU}^{\mu}_{\epsilon}(\{(f,g)\}) = \mathsf{CC}^{\mu}_{\epsilon}(g)$. Furthermore the uncertain communication complexity is monotone, i.e., if $\mathcal{F} \subseteq \mathcal{F}'$ then $\mathsf{CCU}^{\mu}_{\epsilon}(\mathcal{F}) \leq \mathsf{CCU}^{\mu}_{\epsilon}(\mathcal{F}')$. So we conclude that $\mathsf{CCU}^{\mu}_{\epsilon}(\mathcal{F}) \geq \max_{\{g \mid \exists f \text{ s.t. } (f,g)\in\mathcal{F}\}}\{\mathsf{CC}^{\mu}_{\epsilon}(g)\}$.

In the rest of the paper we attempt to identify a setting under which the above lower bound can be matched. If the set of functions $\Gamma(g) = \{f \mid (f,g) \in \mathcal{F}\}$ is not sufficiently informative about g, then it seems hard to conceive settings where Alice can do non-trivially well. We thus pick a simple, natural restriction on $\Gamma(g)$, namely, that it contains functions that are close to g (in δ_μ distance). This leads us to our main target classes. For parameters $k, \epsilon, \delta > 0$ define the set of pairs of functions

$$\mathcal{F}_{k,\epsilon,\delta} \triangleq \{(f,g) \mid \delta_\mu(f,g) \leq \delta \;\&\; \mathsf{CC}^{\mu}_{\epsilon}(f), \mathsf{CC}^{\mu}_{\epsilon}(g) \leq k\}$$

and

$$\mathsf{ow}\mathcal{F}_{k,\epsilon,\delta} \triangleq \{(f,g) \mid \delta_\mu(f,g) \leq \delta \;\&\; \mathsf{owCC}^{\mu}_{\epsilon}(f), \mathsf{owCC}^{\mu}_{\epsilon}(g) \leq k\}.$$

In words, $\mathcal{F}_{k,\epsilon,\delta}$ (resp. $\mathsf{ow}\mathcal{F}_{k,\epsilon,\delta}$) considers all possible functions g with communication complexity (resp. one-way communication complexity) at most k with Alice being roughly under all possible uncertainties within distance δ of Bob.[1]

It is clear that $\mathsf{CCU}^{\mu}_{\epsilon}(\mathsf{ow}\mathcal{F}_{k,\epsilon,\delta}) \geq k$. Our main theorem, Theorem 4.1, gives a comparable upper bound on this quantity (upto a constant factor increase in the error and communication complexity). Before that, in Theorem 3.2 we show that a naive strategy that attempts to reduce the uncertain communication problem to a "function agreement problem" (where Alice and Bob agree on a function q that is close to f and g and then use a protocol for q) cannot work. Achieving a similar understanding of $\mathsf{CCU}^{\mu}_{\epsilon}(\mathcal{F}_{k,\epsilon,\delta})$ is a natural and interesting open question.

3 Hardness of Contextual Agreement

In this section, we show even if both f and g have small one-way distributional communication complexity on some distribution μ, agreeing upon a q such that $\delta_\mu(q,f)$ is small takes communication that is roughly the size of the bit representation of f (which is exponential in the size of the input). Thus, agreeing on q before simulating a protocol for q is exponentially costlier than even the trivial protocol where Alice sends her input x to Bob. Formally, we consider the following communication problem:

Definition 3.1 ($\textsc{Agree}_{\delta,\gamma}(\mathcal{F})$)**.** *For a family of pairs of functions $\mathcal{F} \subseteq \{f\colon X \times Y \to \{0,1\}\}^2$ the $\mathcal{F}$-agreement problem with parameters $\delta, \gamma \geq 0$ is the communication problem where Alice gets f and Bob gets g such that $(f,g) \in \mathcal{F}$ and their goal is to for Alice to output q_A and Bob to output q_B such that $\delta(q_A,f), \delta(q_B,g) \leq \delta$ and $\Pr[q_A = q_B] \geq \gamma$.*

Abusing notation somewhat we will use $\textsc{Agree}_{\delta,\gamma}(\mathcal{D})$ to denote the distributional problem where $\mathcal{D}$ is a distribution on $\{f\colon X \times Y \to \{0,1\}\}^2$ and the goal now is to get agreement with probability δ over the randomness of the protocol and the input.

[1] For the sake of symmetry we insist that $\mathsf{CC}^{\mu}_{\epsilon}(f) \leq k$ (resp. $\mathsf{owCC}^{\mu}_{\epsilon}(f) \leq k$). We need not have insisted on it but the other conditions anyhow imply that $\mathsf{CC}^{\mu}_{\epsilon+\delta}(f) \leq k$ (resp. $\mathsf{owCC}^{\mu}_{\epsilon+\delta}(f) \leq k$, respectively) so we decided to include this stronger condition for aesthetic reasons.

If the agreement problem could be solved with low communication for the family $\mathcal{F}_{k,\delta,\epsilon}$ as defined at the end of Section 2, then this would turn into a natural protocol for $\mathsf{CCU}(\mathcal{F}_{k,\epsilon',\delta'})$ for some positive ϵ' and δ' as well. Unfortunately, our theorem below proves that agreement is a huge overkill.

Theorem 3.2. *For every $\delta, \delta_2 > 0$ there exists $\alpha > 0$ and a family $\mathcal{F} \subseteq \mathcal{F}_{0,0,\delta}$ such that for every $\gamma > 0$ it holds that $\mathsf{CC}(\textsc{Agree}_{\delta_2,\gamma}(\mathcal{F})) \geq \alpha|Y| - \log(1/\gamma)$.*

In words the theorem says that there is a family of pairs of functions supported on functions of *zero* communication complexity (with zero error) for which agreement takes communication polynomial in the size of the domain of the functions. Note that this is exponentially larger than the trivial communication complexity for any function g, which is at most $\min\{1+\log|Y|, \log|X|\}$.

We stress that while an agreement lower bound for zero communication functions may feel a lower bound for a toy problem, a lower bound for this setting is inherent in any separation between agreement complexity for $\mathcal{F}$ and communication complexity with uncertainty for $\mathcal{F}$. To see this note that given any input to the $\mathsf{CCU}(\mathcal{F})$ problem, Alice and Bob can execute any protocol for $\mathsf{CCU}(\mathcal{F})$ pinning down value of the function to be computed with high probability and low communication. If one considers the remaining challenge to agreement, it comes from a zero communication problem.

Our proof of Theorem 3.2 uses a lower bound on the communication complexity of *agreement distillation (with imperfectly shared randomness)* problem defined in [CGMS14], who in turn rely on a lower bound for randomness extraction from correlated sources due to Bogdanov and Mossel [BM11].

We describe their problem below and the result we use. We note that their context is slightly different and our description below is a reformulation. First we define the notion of ρ-perturbed sequences of bits. A pair of bits (a, b) is said to be ρ-perturbed uniform bits if a is uniform over $\{0,1\}$ and $b = a$ with probability $1-\rho$ and $b \neq a$ with probability ρ. A pair of sequence of bits (r, s) is said to be ρ-perturbed if $r = (r_1, \ldots, r_n)$ and $s = (s_1, \ldots, s_n)$ and each coordinate pair (r_i, s_i) is ρ-perturbed uniform pair drawn independently of all other pairs. For a random variable W, we let its min-entropy, denoted $H_\infty(w) = \min_{w \in \mathsf{supp}(W)}\{-\log(Pr[W = w]\}$.

Definition 3.3 ($\textsc{Agreement-Distillation}^k_{\gamma,\rho}$)**.** *In this problem, Alice and Bob get as inputs r and s where (r, s) form a ρ-perturbed sequence of bits. Their goal is to communicate deterministically and produce as outputs w_A (Alice's output) and w_B (Bob's output) with the following properties: (i) $H_\infty(w_A), H_\infty(w_B) \geq k$ and (ii) $\Pr_{(r,s)}[w_A = w_B] \geq \gamma$.*

Lemma 3.4 ([CGMS14, Theorem 2.6])**.** *For every $\rho > 0$ there exists $\epsilon > 0$ such that for every k and γ it holds that every deterministic protocol Π that computes $\textsc{Agree}^k_{\gamma,\rho}$ has communication complexity at least $\epsilon k - \log 1/\gamma$.*

We note that while the agreement distillation problem is very similar to our agreement problem, there are syntactic differences — we are considering pairs of functions with low communication complexity, whereas the agreement-distillation problem considers arbitrary random sequences. Also our output criterion is proximity to the input functions, whereas in the agreement-distillation problem we need to produce high-entropy outputs. Finally, we want a lower bound for our agreement problem when Alice and Bob are allowed to share perfect randomness while the agreement-distillation bound only holds for deterministic protocols. Nevertheless, we are able to reduce to their setting quite easily as we will see shortly.

Our proof of Theorem 3.2 uses the standard Chernoff-Hoeffding tail inequality on random variables that we include below. Denote $\exp(x) \triangleq e^x$, where e is the base of the natural logarithm.

Proposition 3.5 (Chernoff bound)**.** *Let $X = \sum_{i=1}^{n} X_i$ be a sum of identically distributed independent random variables $X_1, \ldots, X_n \in \mathbb{B}$. Let $\mu = \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i]$. It holds that for $\delta \in (0,1)$,*

$$\Pr[X < (1-\delta)\mu] \leq \exp\left(-\delta^2\mu/2\right)$$

and

$$\Pr[X > (1+\delta)\mu] \leq \exp\left(-\delta^2\mu/3\right).$$

Proof of Theorem 3.2. We prove the theorem for $\alpha < \delta/6$, in which case we may assume $\gamma > \exp(-\delta|Y|/6)$ since otherwise the right hand side is non-positive.

Let $\mathcal{F}_B$ denote the set of functions that depend only on Bob's inputs, i.e., $f \in \mathcal{F}_B$ if there exists $f' \colon Y \to \{0,1\}$ such that $f(x,y) = f'(y)$ for all x, y. Our family $\mathcal{F}$ will be a subset of $\mathcal{F}_B \times \mathcal{F}_B$, the subset that contains functions that are at most $\delta|Y|$ apart.

$$\mathcal{F} \triangleq \{(f,g) \in \mathcal{F}_B \times \mathcal{F}_B \mid \delta(f,g) \leq \delta\}.$$

It is clear that communication complexity of every function in the support of $\mathcal{F}$ is zero, with zero error (Bob can compute it on his own) and so $\mathcal{F} \subseteq \mathcal{F}_{0,0,\delta}$. So it remains to prove a lower bound on $\mathsf{CC}(\textsc{Agree}_{\delta_2,\gamma}(\mathcal{F}))$.

We prove our lower bound by picking a distribution $\mathcal{D}_\rho$ supportedly mostly on $\mathcal{F}$ and by giving a lower bound on $\mathsf{CC}(\textsc{Agree}_{\delta_2,\gamma}(\mathcal{D}_\rho))$. Let $\rho = \delta/2$. The distribution $\mathcal{D}_\rho$ is a simple one: It samples (f,g) as follows. f is drawn uniformly at random from $\mathcal{F}_B$. g is then chosen to be a "ρ-perturbation" of f, namely for every $y \in Y$, $g'(y)$ is chosen to be equal to $f(x,y)$ with probability $1-\rho$ and $1-f(x,y)$ with probability ρ. For every $x \in X$ we now set $g(x,y) = g'(x,y)$.

By the Chernoff bound (see Proposition 3.5) we have that $\Pr_{(f,g)\sim\mathcal{D}_\rho}[\delta(f,g) > \delta] = \exp(-\rho|Y|/3) \leq \gamma$. So with overwhelmingly high probability $\mathcal{D}_\rho$ draws elements from $\mathcal{F}$. In particular, if some protocol solves $\textsc{Agree}_{\delta_2,\gamma}(\mathcal{F})$, then it would also solve $\textsc{Agree}_{\delta_2,2\gamma}(\mathcal{D}_\rho)$.

We thus need to show a lower bound on the communication complexity of $\textsc{Agree}_{\delta_2,2\gamma}(\mathcal{D}_\rho)$. We now note that since this is a distributional problem, by the Yao's min-max principle if there is randomized protocol to solve $\textsc{Agree}_{\delta_2,2\gamma}(\mathcal{D}_\rho)$ with C bits of communication, then there is also a deterministic protocol for the same problem with same complexity. Thus, it suffices to lower bound the deterministic communication complexity of $\textsc{Agree}_{\delta_2,2\gamma}(\mathcal{D}_\rho)$. Claim 3.6 shows any such protocol gives a deterministic protocol for Agreement-Distillation with $k = \Omega_{\delta_2}(|Y|)$. Combining this with Lemma 3.4 gives us the desired lower bound on $\mathsf{CC}(\textsc{Agree}_{\delta_2,2\gamma}(\mathcal{D}_\rho))$ and hence on $\mathsf{CC}(\textsc{Agree}_{\delta_2,\gamma}(\mathcal{F}))$. □

Claim 3.6. *Every protocol for $\textsc{Agree}_{\delta_2,\gamma}(\mathcal{D}_\rho)$ is also a protocol for $\textsc{Agreement-Distillation}^{k}_{\gamma,\rho}$ for $k = (1-h(\delta_2))|Y|$, where $h(\cdot)$ is the binary entropy function given by $h(x) = -x\log x - (1-x)\log(1-x)$.*

Proof. Suppose Alice and Bob are trying to solve $\textsc{Agreement-Distillation}^{k}_{\gamma,\rho}$. They can sample ρ-pertubed strings $(r,s) \in \{0,1\}^{|Y|}$ and interpret them as functions $f', g' : Y \to \{0,1\}$ or equivalently as functions $(f,g) \sim \mathcal{D}_\rho$. They can now simulate the protocol for $\textsc{Agree}_{\delta,\gamma}(f,g)$ and output q_A and q_B. By definition of Agree, we have $q_A = q_B$ with probability at least γ. So it suffices

to show that $H_\infty(q_A), H_\infty(q_B) \geq k$. But this is obvious since any function q_A is output only if $\delta(f, q_A) \leq \delta_2$ and we have $\{f \mid \delta(f, q_A) \leq \delta\} \leq 2^{h(\delta_2)|Y|}$. Since the probability of sampling f for any f is at most $2^{-|Y|}$, we have that the probability of outputting q_A for any q_A is at most $2^{-(1-h(\delta_2))|Y|}$. In other words $H_\infty(q_A) \geq (1 - h(\delta_2))|Y|$. Similarly, we can lower bound $H_\infty(q_B)$ and thus we have that the outputs of the protocol for Agree solve Agreement-Distillation with $k = (1 - h(\delta_2))|Y|$. □

4 One-way Communication Complexity with Contextual Uncertainty

In this section, we describe our main result that allows us to transform any one-way protocol in the standard distributional complexity framework to a one-way protocol under contextual uncertainty with only a small multiplicative overhead in the communication complexity, when Alice and Bob's inputs are sampled from a product distribution.

Let $\mu = \mu_X \times \mu_Y$ be a distribution over an input space $X \times Y$. For parameters $k, \epsilon, \delta > 0$ recall the set of pairs of functions defined previously

$$\mathsf{ow}\mathcal{F}^{\mu}_{k,\epsilon,\delta} \triangleq \{(f, g) | \delta_\mu(f, g) \leq \delta \ \&\ \mathsf{owCC}^{\mu}_{\epsilon}(f), \mathsf{owCC}^{\mu}_{\epsilon}(g) \leq k\}.$$

In this section, we show that for any $\theta > 0$ it holds that $\mathsf{owCCU}_{\epsilon+2\delta+\theta}(\mathsf{ow}\mathcal{F}^{\mu}_{k,\epsilon,\delta}) = O(k/\theta^2)$.

Theorem 4.1. *There exists a constant $c < \infty$ such that for every pair of finite sets X and Y, for every product distribution $\mu = \mu_X \times \mu_Y$ over $X \times Y$ and for every $\theta > 0$ it holds that $\mathsf{owCCU}^{\mu}_{\epsilon+2\delta+\theta}(\mathsf{ow}\mathcal{F}^{\mu}_{k,\epsilon,\delta}) \leq ck/\theta^2$.*

Proof overview and some notation. Let us first introduce convenient notation that will help us describe our idea and later the actual protocol in the distributional communication complexity model. For any function $s\colon X \times Y \to \{0, 1\}$ and $x \in X$, let us define the *restriction* of s to x to be the function $s_x\colon Y \to \{0, 1\}$ given by $s_x(y) = s(x, y)$ for $y \in Y$.

Our main idea is that Alice, given her inputs (f, x) can determine f_x, and will try to describe it to Bob. For most x, f_x will be close (in δ_{μ_Y}-distance) to the function g_x. Bob will try to use the (as yet unspecified) description given by Alice to determine some function B close to g_x. If he succeeds, he can output $B(y)$ and this should work with high probability over y (based on the definition of closeness).

It remains to explain how Alice will describe f_x, and how Bob will determine some function B close to g_x based on Alice's description. For the first part we follow a natural idea. We let Alice and Bob use shared randomness[2] to sample $y_1, \ldots, y_m$ where the y_i's are drawn independently with $y_i \sim \mu_Y$, and m is a parameter to be chosen later. Alice's description of f_x will simply be $(f_x(y_1), \ldots, f_x(y_m)) \in \{0, 1\}^m$. So the length of the communication is m bits and we need to show that setting m to be roughly $O(k)$ suffices. Before we explain this, we first need to specify what Bob does with Alice's message.

As a first cut, let us consider the following natural strategy: Bob simply picks an $\tilde{x} \in X$ such that $g_{\tilde{x}}$ is close to f_x on $z_1, \ldots, z_m$, and sets $B = g_{\tilde{x}}$. It is clear that if $\tilde{x} = x$ then $B = g_{\tilde{x}} = g_x$ so

[2] As mentioned earlier, shared randomness is not necessary in our setting, but clarifies the explanation so we use it.

for every $y \in \mu_Y$ we have $B(y) = g_x(y)$. Indeed if $\tilde{x}$ is such that $g_{\tilde{x}}$ is close to $g_x \approx f_x$ then again we are ok since $B(y)$ will now equal $g_x(y)$ with high probability. It remains to deal with $\tilde{x}$ such that $g_{\tilde{x}}$ is far from g_x. If we fix any such $\tilde{x}$ and then choose $y_1, \ldots, y_m$ afterwards then again we are fine. Since g_x is close to f_x, $g_{\tilde{x}}$ is far also from f_x and so randomly chosen $y_1, \ldots, y_m$ should reveal this. However this can not deal with all possible $\tilde{x}$ — to use a naive union bound over all possible $\tilde{x} \in X$ would require a failure probability of $(1/|X|)$ which would require $m \approx \log |X|$. Indeed smaller m should not suffice since we haven't yet used the fact that $\mathsf{CC}^{\mu}_{\epsilon}(g) \leq k$ — but we do so next.

Suppose Π is a one-way protocol with k bits of communication. Then note that Alice's message partitions X into 2^k sets, one corresponding to each message. Our modified strategy for Bob is to pick a representative x from each partition, and then to set $B = g_{\tilde{x}}$ for an $\tilde{x}$ among the representatives for which $g_{\tilde{x}}$ and f are close on the samples $y_1, \ldots, y_m$. A simple analysis reveals that g_x's within a partition are $O(\epsilon)$-close, so if we pick $\tilde{x}$ to be the representative of the partition containing x then $g_{\tilde{x}}$ and f_x will be close on the sampled points. For the other representatives, once again if $g_{\tilde{x}}$ is close to g_x for some representative, then we are happy since then $g_{\tilde{x}}(y)$ will equal $g_x(y)$ with high probability. For a representative $\tilde{x}$ such that $g_{\tilde{x}}$ is far from $g_x \approx f_x$ we can proceed as before; and now the union bound works out since the number of representatives is only 2^k.

In what follows we formalize the argument above.

Proof of Theorem 4.1. Recall that Alice's input in the contextual setting is (f, g) and Bob's input is (g, y) where $(f, g) \in \mathsf{ow}\mathcal{F}^{\mu}_{k,\epsilon,\delta}$ and $(x, y) \sim \mu$.

Let Π be the one-way protocol for g (in the standard setting) that shows $\mathsf{owCC}^{\mu}_{\epsilon}(g) \leq k$. Note that Π can be described by an integer $L \leq 2^k$ and functions $\pi : X \to [L]$ and $\{B_i : Y \to \{0,1\}\}_{i \in [L]}$, such that Alice's message on input x is $\pi(x)$ and Bob's output on message i from Alice and input y is $B_i(y)$. We use this notation below.

In what follows we will use a parameter $m = \Theta(k/\theta^2)$ chosen such that $2^k \cdot \exp(-\theta^2 m/27) \leq \theta/3$.

The protocol. Algorithm 1 describes the protocol we employ in the contextual setting. Roughly speaking, the protocol works as follows. First, Alice and Bob sample random points $y_1, y_2 \ldots, y_m \in Y$. The Alice sends the sequence $(f_x(y_1), \ldots, f_x(y_m))$ to Bob. Bob enumerates $i \in [L]$ and counts the fraction of $z \in \{y_1, \ldots, y_m\}$ for which $B_i(z) \neq f_x(z)$. For i such that this fraction is minimized, Bob outputs $B_i(y)$ and halts.

Algorithm 1: The protocol that handles contextual uncertainty

The setting: Let μ be a probability distribution over a message space $X \times Y$. Alice and Bob are given functions f and g, and inputs x and y, respectively, where $(f, g) \in \mathsf{ow}\mathcal{F}^{\mu}_{k,\epsilon,\delta}$ and $(x, y) \sim \mu$.

The protocol:

1. Alice and Bob parse the common random string as a sample of m values $Z = \{y_1, y_2, \ldots, y_m\} \subseteq Y$ each sampled (independently) according to μ_Y.
2. Alice sends $\{f_x(y_i)\}_{i \in [m]}$ to Bob.
3. For every $i \in [L]$, Bob computes $\mathsf{err}_i \triangleq \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}(B_i(y_j) \neq f_x(y_j))$. Let $i_{\min} = \mathrm{argmin}_{i \in [L]}\{\mathsf{err}_i\}$. Bob outputs $B_{i_{\min}}(y)$ and halts.

Observe that the amount of communication in the protocol is exactly $m = \Theta(k/\theta^2)$ bits, as promised. Next, we analyze the correctness of the protocol.

Analysis. We start with some notation: For $x \in X$, let $\delta_x \triangleq \delta_{\mu_Y}(f_x, g_x)$ and let $\epsilon_x \triangleq \delta_{\mu_Y}(g_x, B_{\pi(x)})$. Note by definition that $\delta = \mathbb{E}_{x\sim\mu_X}[\delta_x]$ and $\epsilon = \mathbb{E}_{x\sim\mu_X}[\epsilon_x]$. For $i \in [L]$, let $\gamma_{i,x} \triangleq \delta_{\mu_Y}(f_x, B_i)$. Note that $\gamma_{\pi(x),x} = \delta_{\mu_Y}(f_x, B_{\pi(x)}) \leq \delta_x + \epsilon_x$.

In what follows we will analyze the probability that $B_{i_{\min}}(y) \neq g(x, y)$ by analyzing the estimates err_i and the index $i_{\min}$ computed in the protocol above. Note that $\mathsf{err}_i = \mathsf{err}_i(x)$ computed above attempts to estimate $\gamma_{i,x}$ and both err_i and $i_{\min}$ are functions of x.

By a simple application of the Chernoff bound (see Proposition 3.5) we have for every x and $i \in [L]$

$$\Pr_{y_1,\ldots,y_m}[|\gamma_{i,x} - \mathsf{err}_i| > \theta/3] \leq \exp(-\theta^2 m/27).$$

By a union bound we have for every $x \in X$,

$$\Pr_{y_1,\ldots,y_m}[\exists i \in [L] \text{ s.t. } |\gamma_{i,x} - \mathsf{err}_i| > \theta/3] \leq L \cdot \exp(-\theta^2 m/27) \leq \theta/3.$$

Now assume $\forall i \in [L]$, we have $|\gamma_{i,x} - \mathsf{err}_i| \leq \theta/3$ (which we refer to below as the "Good Event"). Then, for $i_{\min}$ we have

$$\begin{aligned}
\gamma_{i_{\min},x} &\leq \mathsf{err}_{i_{\min}} + \theta/3 \text{ (since we assumed the Good Event)} \\
&\leq \mathsf{err}_{\pi(x)} + \theta/3 \text{ (By definition of } i_{\min}) \\
&\leq \gamma_{\pi(x),x} + 2\theta/3 \text{ (Again because of the Good Event)} \\
&\leq \delta_x + \epsilon_x + 2\theta/3.
\end{aligned}$$

Thus we have assuming the Good Event that Bob's output $B_{i_{\min}}(y)$ satisfies $\Pr_{x,y}[B_{i_{\min}}(y) \neq f(x, y)] \leq \mathbb{E}_x[\delta_x + \epsilon_x + 2\theta/3] = \delta + \epsilon + 2\theta/3$. Accounting the probability that the good event does not happen, we get that Bob's output disagrees with $f(x, y)$ with probability at most $\epsilon + \delta + \theta$. Finally, since $\delta(f, g) \leq \delta$, we have that Bob's output does not equal $g(x, y)$ (which is the desired output) with probability at most $\epsilon + 2\delta + \theta$. □

References

[BM11] Andrej Bogdanov and Elchanan Mossel. On extracting common random bits from correlated sources. *IEEE Transactions on Information Theory*, 57(10):6351–6355, 2011.

[CGMS14] Clément L. Canonne, Venkatesan Guruswami, Raghu Meka, and Madhu Sudan. Communication with imperfectly shared randomness. *CoRR*, abs/1411.3603, 2014. Extended abstract appears in *Innovations in Theoretical Computer Science, ITCS* 2015.

[GJS12] Oded Goldreich, Brendan Juba, and Madhu Sudan. A theory of goal-oriented communication. *J. ACM*, 59(2):8, 2012.

[HS14] Elad Haramaty and Madhu Sudan. Deterministic compression with uncertain priors. In *Innovations in Theoretical Computer Science, ITCS*, pages 377–386. ACM, 2014.

[JKKS11] Brendan Juba, Adam Tauman Kalai, Sanjeev Khanna, and Madhu Sudan. Compression without a common prior: an information-theoretic justification for ambiguity in language. In *Innovations in Computer Science - ICS*, pages 79–86. Tsinghua University Press, 2011.

[JS08] Brendan Juba and Madhu Sudan. Universal semantic communication I. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 123–132. ACM, 2008.

[JS11] Brendan Juba and Madhu Sudan. Efficient semantic communication via compatible beliefs. In *Innovations in Computer Science - ICS*, pages 22–31. Tsinghua University Press, 2011.

[JW13] Brendan Juba and Ryan Williams. Massive online teaching to bounded learners. In *Innovations in Theoretical Computer Science, ITCS*, pages 1–10. ACM, 2013.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the 11h Annual ACM Symposium on Theory of Computing*, pages 209–213. ACM, 1979.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de