PNAS RESEARCH ARTICLE | CLASSIFICATION

New fundamental limits incorporating logical reasoning into Shannon's information theory

Luis A. Lastras[a], Jonathan Lenchner[a], Barry M. Trager[a], Wojciech Szpankowski[b,e], Mark S. Squillante[a], Chai Wah Wu[a], Ronald Fagin[d], and Alexander Gray[c,b]

This manuscript was compiled on July 15, 2026

A cornerstone of Shannon's famous information theory is the idea of decoupling the meaning of a message from its efficient transmission. In this article, we propose an extension of Shannon's communication model where the sender and receiver are assumed to have reasoning capabilities. In such a setting, we gain new insights by coupling the fields of information theory and mathematical logic. Under the assumption that a message is coming from a stochastic source, Shannon's theory establishes that the fundamental compression limit is the entropy of the source. Imagine, however, that we were not interested in the message itself, but rather, we were focused on the logical conclusions that one could derive from it. In this work, we obtain a closed-form expression for a new fundamental compression limit in the presence of reasoning capabilities. Our new expression is valid under various assumptions about what the sender and receiver know, providing initial answers to key questions such as how much the fundamental limit varies as one narrows or widens the amount of knowledge that is being transferred from sender to receiver, or how, surprisingly, such a fundamental limit remains the same even if the sender is unaware of what it is that a receiver already knows. We also offer practical algorithms that are empirically demonstrated to be significantly more efficient than alternatives that do not account for the existence of reasoning capabilities.

Logical information | Semantic communication | Reasoning

Significance

The value of a bit can range from being extremely consequential to being unimportant, depending on what one can deduce from such a bit. Consider a bit from a reliable sensor that tells a car when to stop versus one from a faulty sensor producing random fluctuations: the value of each depends entirely on what one can deduce from it. Yet today, when designing communication protocols, we do so without accounting for the potential advantages of reasoning capabilities over received bits. We present, for the first time, a theory that accounts for the possibility of reasoning over bits received, which, compared with a classical information-theoretic treatment, demonstrates substantial gains in communication efficiency across multiple tasks, by exploiting previously unappreciated connections between information theory and mathematical logic.

In his famous Lectures on Physics, Feynman imagined a cataclysmic situation in which all scientific knowledge is destroyed, and a sentence carrying the most information in the fewest words needs to be chosen for transmission to future generations of beings (1). His choice for that single sentence — "...all things are made of atoms – little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another." — arguably implies an enormous amount of knowledge for anyone willing to think carefully about it.

The essential point of this example is the power of reasoning and experimentation when seeded with the right critical information. Interestingly, Shannon's information theory intentionally does not account for the presence of computational devices that can reason over received information. This was part of a more general and deliberate decision to ignore semantics, focusing purely on the statistical transmission of symbols rather than their meaning.

This paper proposes to extend that framework (Figure 1) to account for reasoning by the receiver — specifically, logical deduction — which can dramatically amplify the effective content of transmitted information. We show that when senders and receivers can reason, significant efficiency gains over classical Shannon communication are achievable.

Background. In addition to its central role in communications, artificial intelligence (AI) and many other engineering fields, the formal framework of information theory has proved vital for modeling many of the most foundational scientific phenomena across the natural sciences.

Author affiliations: [a]IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA; [b]Purdue University, West Lafayette, IN 47907, USA; [c]Centaur AI Institute, Lincoln, CA 95648, USA; [d]IBM Research-Almaden 95120, USA; [e]Jagiellonian University, Kraków, Poland

L.A.L. contributed the formulation and theorem proofs of logic problems as a variety of information and coding theory problems. J.L. contributed the logical framework and experiments. B.M.T. contributed to algorithms, software, experiments and analysis as well as extensions to finite field algebra. W.S. made contributions to the problem formulation and correctness of the information-theoretic proofs. M.S.S. contributed improvements on the soundness of the mathematical aspects as well as experimental methodology, and extensions to finite field algebra. C.W.W. contributed to the formulation at the intersection of information theory and logic. R.F. contributed to the narrative and provided insights that significantly improved the detailed exposition. A.G. conceived the importance of work connecting these two fields and made contributions to the formulation and positioning of this work relative to the field of logic.

[1]E-mail: lastrasl@us.ibm.com

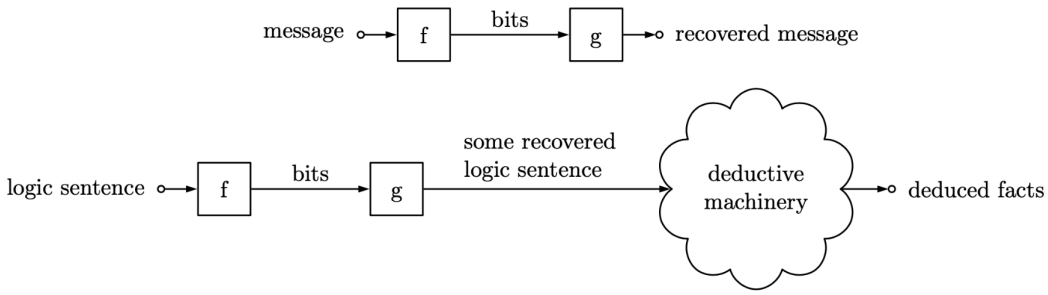


Fig. 1. At the top, Shannon's original digital communication model. At the bottom, a sketch of our proposed variant of Shannon's model.

It has been recognized within physics as a pivotal framework for understanding many core problems that cut across numerous areas of physics (2), including atoms in molecules (3), crystal structure (4), molecular interactions (5), and complex systems (6). It has been particularly essential in quantum computing (7), quantum physics more broadly (8), and in statistical thermodynamics (9, 10), to the extent that it was famously argued by Wheeler that all physical reality can be seen as a manifestation of the more fundamental notion of information (11). Particularly in the age of increasingly rich measurements and datasets, information theory has become an indispensable tool in biology for unlocking the fundamental principles underlying the biochemical signaling at the heart of the functioning of the cell (12–15), as well as advancing understanding in genomics (16–18) and drug response (19). Information-theoretic ideas also help to explain countless other scientific phenomena, from neuroscience (20–22) to human language (23) to art (24) to climate change (25). We expect the deep extensions to information theory described herein to provide correspondingly broad modeling insights across the sciences (26).

The success of notions of semantics-free information is associated with their wide generality. However, reintroduction of semantics has been of recurrent interest (27). In 1952, Carnap and Bar-Hillel (28) utilized the long-standing formalization of semantics as *logic*, which can be traced from Aristotle (logic as syllogism) through to Boole (Boolean logic) and Frege (First-Order Logic), to propose a mathematical conceptualization of a semantic information theory. In his 1953 paper on lattices and information theory (29), Shannon himself argued that entropy "can hardly be said to represent the actual information."

In Carnap and Bar-Hillel's work we find the key idea that if a sentence can be derived from the sentences one already knows, it conveys zero information. Similarly, in (29), Shannon defined the "distance" between two sentences to be zero if each can be derived from the other. In both cases, the more general implications to communication systems were left unexplored. A number of proposals have since been put forward (30–44), broadly addressing one of three areas: the fundamental question of how to define semantic information, the transmission of semantic information, or the learning of models for it. In a subset of this work that has deeper connections to classical information theory (34–40), a common thread is the exploitation of Rate-Distortion theory (45), which Shannon introduced to extend the lossless compression results of (46) to lossy coding (47). In these works, semantics are introduced by asserting that the fidelity criterion in lossy coding could be semantic in nature; they have not, however, explicitly considered reasoning in their frameworks. A different line of investigation (43, 44) starts from Shannon's information lattice paper (29) and then constructs learning methods that are able to obtain explainable rules at different levels of abstraction. In another distinctive line of research, a theory of communication in the presence of goals has been put forth (41, 42).

AI has a long history of incorporating semantic understanding of the world into models, particularly through logic, ever since its origins in the 1950s (48). The rise and current dominance of semantics-free machine learning perhaps mirrors that of information theory, as technological developments have allowed it to achieve great successes but have also sparked discussion about the way ahead beyond current limits. This can be seen in an acceleration of interest in improving upon today's deep learning models (which, in fact, heavily utilize classical information theory) toward those that encode semantics in various ways, as illustrated by Yann LeCun's recent work (49) and the rapid growth of recent work in neuro-symbolic AI (50–52).

Setup. The purpose of this section is to augment the general intuition given by Figure 1 with a formal mathematical framework. We refer the reader to Figure 2(a), which shows a sender (Alice), a receiver (Bob), and sentences $\mathtt{S}$, $\mathtt{Q}$, and $\mathtt{R}$ that play different roles depending on the choice of setup. For now, the reader may assume that these sentences come from some, as yet unspecified, logical language $\mathcal{L}$. The sentences $\mathtt{S}$ and $\mathtt{R}$ represent what Alice and Bob know to be true, respectively, and the query $\mathtt{Q}$ represents a fact that can be deduced starting from $\mathtt{S}$; we say that $\mathtt{Q}$ is provably entailed by $\mathtt{S}$, and write $\mathtt{S} \vdash \mathtt{Q}$. The notion of provable entailment represents the cornerstone we use to bridge between the fields of information theory and mathematical logic. The choice of uppercase notation for these sentences underlines that we think of them as drawn from some distribution; the purpose of this is to be able to talk about expected performance across a class of problems.

Figure 2(a) depicts the deductive machinery from Figure 1 as a diamond. Consider a starting sentence:

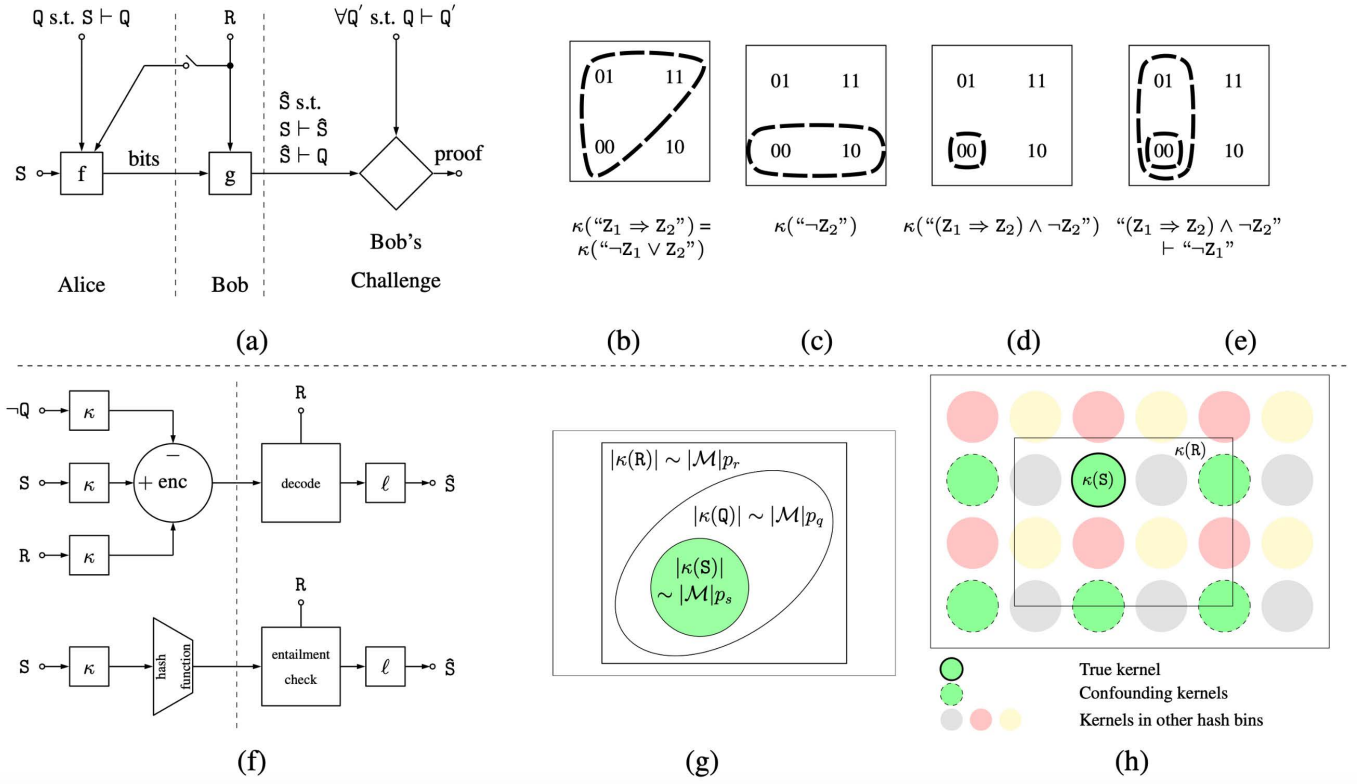


Fig. 2. Communication problem setup and algorithm innovations. (**a**) *General communication diagram* which allows for Bob to possess facts that Alice may or may not know, and which allows for the query Q that Alice is helping Bob prove to be anything from her full knowledge S to a more targeted logic sentence. (**b**)-(**e**) *Kernel diagrams* respectively showing equivalence, negation, conjunction, and provable entailment for a two variable binary propositional logic setting $Z_i \in \{0, 1\}, i = 1, 2$; here $Z_i = 1$ (respectively, $Z_i = 0$) denotes that the ith logic propositional variable is True (respectively, False). (**f**) *Theoretically optimal solution architectures* for the cases where Alice knows and does not know Bob's knowledge, respectively. (**g**) *Probabilistic model for kernels* where the parameters p_s, p_q, p_r are average kernel sizes, illustrated in the case Alice's knowledge implies that of Bob. (**h**) *No need to know* decoding mechanism example where Alice mapped her kernel to a hash bin from 4 possible ones; Bob is able to reconstruct $\kappa(\mathtt{S})$ without Alice knowing R ($p_r < 1$) by rejecting all kernels that do not provably entail R and that do not match the received bin index.

> A force of magnitude F imparts an acceleration on a mass of magnitude m equal to F/m (an expression more conventionally expressed as $F = ma$).

And a query to potentially be proved:

> Given two objects dropped from above the earth's surface at the same height, the time until impact is the same regardless of their mass, disregarding the effect of air friction.

The deductive machinery will either be able to prove the query from the starting sentence, or conclude that the query cannot be proved from the starting logic sentence. Importantly, however, the deductive machinery may have access to additional background facts, such as:

> The force imparted on either mass, m_1 or m_2, of a two-mass system with each mass at a distance r from one another is given by the equation $F = G\frac{m_1 m_2}{r^2}$, where G is a constant.

The issues of (i) what constitutes a valid sentence, given suitable syntactic rules, (ii) how one defines what it means for a sentence to be true or false, and (iii) how one deduces one sentence from another (or from another set of sentences) given a proof system, $\mathcal{P}$, are at the foundations of mathematical logic. In an extended version of this article (53) we carefully construct a bridge between these foundational logical concepts and our theory of semantic communication. The provable entailment operator, $\vdash$, which we utilize informally here, is more formally defined with respect to a particular choice of a proof system, $\mathcal{P}$, in (53).

Now, imagine multiple objects with different masses at some distance from the earth. Each of these objects has a somewhat different force imparted on it based on its mass and the acceleration it experiences and additionally has its own time until impact. Each earth-object system is in this sense a "model" of the various physics sentences (the starting sentence, the query to be proved, and the background knowledge).

This notion of a model carries over to the formal logical theory we develop in (53) where a model of a sentence is assumed to come from a set $\mathcal{M}$ of possible models. A logic sentence is thus intrinsically associated with those models within $\mathcal{M}$ that make the sentence true. The subset of models of a sentence $\mathtt{S} \in \mathcal{L}$ is called the *kernel* of such a sentence and is denoted by $\kappa(\mathtt{S})$. Notice that two different sentences from the language could be associated with exactly the same set of models, which makes those sentences logically equivalent.

This last observation is the key foundation for our contributions. Our work places very few restrictions on the choice of logic, language $\mathcal{L}$ and set $\mathcal{M}$ of models, with an important limitation that the latter currently needs to be finite. This includes propositional and first order logic, but applies to additional logics and languages that

we have yet to investigate. The limitation to a finite $\mathcal{M}$ was introduced to obtain the simplest possible non-trivial theory combining information theory and logic; we intend to remove such a limitation in subsequent treatments of this work.

For illustrative purposes, in Figure 2 we consider propositional logic with 2 variables $\mathtt{Z}_1, \mathtt{Z}_2 \in \{0, 1\}$, where $\mathtt{Z}_\mathtt{i} = 0$ if the ith proposition is false and $\mathtt{Z}_\mathtt{i} = 1$ if true. A set of models then coincides with a set of possible truth assignments to these variables. We depict three such kernels in Figures 2(b)-(d). Figure 2(b) also illustrates how two different sentences that are logically equivalent have the same kernel, and Figure 2(d) shows how the kernel of a conjunction of sentences is the intersection of the individual kernels. Figure 2(e) illustrates a concept from mathematical logic that we use extensively in our article: a logic sentence provably entails another logic sentence if and only if the kernel of the former is a subset of the kernel of the latter. The concept of a kernel in the context of a semantic information theory traces its roots back to Carnap and Bar-Hillel, who defined a closely related concept using the word "range".

Initial meeting. Alice and Bob meet ahead of time, and agree that the goal is for Bob to prove the truth of a logic sentence $\mathtt{Q} \in \mathcal{L}$ using information that Alice will provide, employing a pre-agreed upon encoding. The sentences $\mathtt{S}, \mathtt{Q}, \mathtt{R}$ are not known at the time of this meeting, and will be revealed later to the relevant parties.

Correlated world observations. After this meeting, Alice and Bob go their own ways; Alice, the sender, obtains knowledge about the world summarized in a logic sentence $\mathtt{S} \in \mathcal{L}$ whereas Bob, the receiver, obtains $\mathtt{R} \in \mathcal{L}$. We consider both settings where Alice knows and does not know $\mathtt{R}$; we do assume that $\mathtt{S} \vdash \mathtt{R}$ except in an interesting incorrect information scenario discussed later. This provable entailment assumption roughly corresponds to a situation where Alice has made a superset of the observations about the world that Bob has made (but may not know exactly what subset Bob has), and where each has independently summarized their observations as logic sentences. The query $\mathtt{Q}$ is also revealed to Alice at this time, where we assume that $\mathtt{S} \vdash \mathtt{Q}$; in words, Alice will help Bob prove something that she can prove with her knowledge. In the case Alice does not know $\mathtt{R}$, we only present the case where Alice is equipping Bob to prove all she can ($\mathtt{Q} = \mathtt{S}$). In the case Alice does know $\mathtt{R}$, we assume that $\mathtt{Q} \vdash \mathtt{R}$. All three of $\mathtt{S}, \mathtt{Q}, \mathtt{R}$ are regarded as stochastic, with the general assumptions on their probability distribution to be described shortly; this is done so as to be able to talk about average performance.

Communication. In the case Alice knows $\mathtt{R}$, she uses an encoder $f(\mathtt{S}, \mathtt{Q}, \mathtt{R})$ to generate bits that are transmitted to Bob. In turn, Bob uses a decoder $g(\mathtt{R}, f(\mathtt{S}, \mathtt{Q}, \mathtt{R}))$ to obtain $\hat{\mathtt{S}}$, which represents Bob's updated logic sentence. In all cases, we assume that Bob ends up with knowledge consistent with that of Alice, written mathematically as $\mathtt{S} \vdash \hat{\mathtt{S}}$ (but the reverse provable entailment may not always hold). In the more challenging setting where Alice does not know $\mathtt{R}$ (and we restrict $\mathtt{Q} = \mathtt{S}$), we allow for multiple communication rounds in which Alice and Bob exchange turns sending information, culminating again with Bob's updated logic sentence $\hat{\mathtt{S}}$. In either case, we use exactly the same figure of merit, defined more precisely below: the number of expected bits exchanged between Alice and Bob, and normalized by the number of models, where the expectation is with respect to $\mathtt{S}, \mathtt{Q}, \mathtt{R}$ for some distribution; once again, the general assumptions on this distribution will be described shortly. Note that the normalization by $|\mathcal{M}|$ is where we most overtly make use of the assumption of the finiteness of $\mathcal{M}$; extensions to the infinite setting will obviously not rely on this specific type of normalization.

This setup has some similarities with Yao's communication complexity (54), with one difference being that we do not identify ahead of time a function both Alice and Bob want to compute. The information-theoretic counterpart of Yao's model, studied by Orlitsky and Roche (55), characterizes the minimum rate at which a sender must communicate so that a receiver can compute a function of their joint data; our setting differs in that the receiver's goal is not to evaluate a pre-specified function but to construct a proof of a logical query.

Challenge and deduction. After the communication takes place, Bob is challenged with any sentence $\mathtt{Q}'$ that can be proven by $\mathtt{Q}$ (including possibly $\mathtt{Q}$ itself), and Bob is able to produce a proof for that query using the logic sentences $\hat{\mathtt{S}}$ and $\mathtt{R}$; mathematically, for the system to have succeeded, it must be the case that $\hat{\mathtt{S}} \vdash \mathtt{Q}$.

Probabilistic model and figure of merit. All of our results are phrased in terms of expectations over various stochastic quantities. The independent variables are normalized expectations of kernel sizes; specifically, $p_s = E[|\kappa(\mathtt{S})|]/|\mathcal{M}|$, $p_q = E[|\kappa(\mathtt{Q})|]/|\mathcal{M}|$, and $p_r = E[|\kappa(\mathtt{R})|]/|\mathcal{M}|$, where $0 < p_s \leq p_q < p_r \leq 1$. For illustration purposes, in the setting of propositional logic with two variables, the normalized sizes of the kernels in Figure 2(b),(c),(d) are 3/4, 1/2 and 1/4, respectively.

The figure of merit for our results is the minimum number of expected bits exchanged, normalized by $|\mathcal{M}|$, defined as

$$\min_{f,g} E[\mathbf{len}(f(\mathtt{S}, \mathtt{Q}, \mathtt{R}))]/|\mathcal{M}|, \qquad [1]$$

where $\mathbf{len}(\cdot)$ returns the length of the binary string that is passed to it, and the minimization is over all possible encoder and decoder pairs f, g that satisfy the provable entailment condition $g(\mathtt{R}, f(\mathtt{S}, \mathtt{Q}, \mathtt{R})) \vdash \mathtt{Q}$ in the case Alice knows $\mathtt{R}$. When Alice does not know $\mathtt{R}$, the setup is more complex; this is partially treated in the subsection *An algorithm for the case that Alice does not know what it is that Bob knows* of our Methods section, with additional details in the supporting information (SI) appendix, and fully described in (53).

We will present a theoretical result in the form of upper and lower bounds on Eq. (1) that agree with each other asymptotically as $|\mathcal{M}|$ goes to infinity; for this reason, we refer to our result as the fundamental limit. Our upper bounds are applicable to any possible distribution over $\mathtt{S}, \mathtt{Q}, \mathtt{R}$ as long as the provable entailment conditions described earlier are met; these are illustrated geometrically in Figure 2(g).

For our lower bounds, we use a stronger assumption that can be roughly described as one where the elements of a kernel are chosen independent and identically distributed (i.i.d.) with some probability, from the elements of a subsuming kernel, resulting in the same overall expected

sizes as before; in this case, p_s, p_q and p_r can be more readily interpreted as probabilities. The precise definitions can be found in the *Conditions for the validity of the lower bound of Theorem 1* subsection of the Methods section. In the case Alice does not know Bob's R, the model by assumption is simpler: only S and R are relevant; we thus set $\mathtt{Q} = \mathtt{S}$ and $p_q = p_s$.

Overview of results. All of our results are phrased in terms of a type of scaled conditional entropy that we call the *logical semantic entropy*, denoted by Λ, which increases monotonically in each variable:

$$\Lambda(a, b) = a \log_2 \left(\frac{a+b}{a} \right) + b \log_2 \left(\frac{a+b}{b} \right).$$

The discovery of the role of this function in problems of efficient communication assuming reasoning capabilities is a key part of our contribution. We note that the same functional form appears independently in recent work on constrained entropy maximization (56, 57), where $\Lambda(a, b)$ governs the entropy cost of merging two probability atoms of sizes a and b, and bounds the approximation gap for the NP-hard problem of entropy-bounded information maximization. In both their setting and ours, Λ emerges as the fundamental cost of a binary separation within a weighted structure, suggesting a deeper role for this function in information theory.

We are now in position to state our main result, to be interpreted in the context of the above setup.

Theorem 1 *For any distribution over* $(\mathtt{S}, \mathtt{Q}, \mathtt{R})$ *meeting the provable entailment conditions* $\mathtt{S} \vdash \mathtt{Q}$ *and* $\mathtt{Q} \vdash \mathtt{R}$*, if the corresponding kernels have expected normalized sizes* p_s, p_q, p_r*, respectively, then for the case Alice knows what* R *is, an algorithm exists with normalized expected cost in bits that is upper bounded by* $\Lambda(p_s, p_r - p_q) + O(|\mathcal{M}|^{-1} \log |\mathcal{M}|)$*. Under additional constraints (see the Methods section), the normalized expected cost of any such algorithm is lower bounded by* $\Lambda(p_s, p_r - p_q)$*. In the case Alice does not know what* R *is, under the additional assumption that* $\mathtt{Q} = \mathtt{S}$ *and thus* $p_q = p_s$*, the same conclusions hold, where the figure of merit accounts for all communication in both directions.*

A sketch of the proof of this Theorem is given in the SI appendix, with the full proof provided in (53). We reiterate that the class of algorithms considered for its lower bound is large, as the result shares similar proof techniques with its corresponding classical information-theoretic cousins. Of note as well is that the upper and lower bounds match up to the approximation dependent on the $|\mathcal{M}|$ parameter, and that the same expression Λ is being used for either of the two settings considered in this article. We also note to the reader that the result above already contains significant advantages with respect to communication protocols that do not rely on reasoning; this will be argued in the remainder of this article.

Solution architecture. In Figure 2(f), we illustrate two architectures that achieve the upper bound in this theorem for the cases where Alice knows and does not know R, respectively. In both cases, we first map the sender's logic sentence S to its kernel $\kappa(\mathtt{S})$. When Alice does know R and the query Q may be more targeted (still respecting the assumption $\kappa(\mathtt{S}) \subseteq \kappa(\mathtt{Q})$), the key technique involves approximating the kernel of Q with another kernel that has two properties: a) it is included in $\kappa(\mathtt{Q})$, and b) it includes $\kappa(\mathtt{S})$; such an approximating kernel is chosen from a carefully crafted list of kernels and its index is transmitted to Bob. This procedure is illustrated as the circle labeled 'enc'. When Alice does not know R, the key idea is compressing Alice's kernel by sending only the index of a bin (hash bin) to which the kernel is mapped by hashing. Bob is able to recover the kernel by choosing the one from the hash bin that provably entails R, thus eliminating confounding ones (Figure 2(h)). To ensure such hash collisions occur with low probability, while maintaining very good compression, the hash bins need to be sized "just right". This is attained by multiple preliminary communication rounds in which sender and receiver communicate their kernel sizes. The small probability event where this protocol runs into difficulties is addressed with one optional final round in which the sender kernel is transmitted with a simple encoding. This small probability can, in effect, be made vanishingly small as $|\mathcal{M}|$ grows. The process ends by using a function ℓ that is able to translate a kernel back to an exemplary logic sentence $\hat{\mathtt{S}}$. Arguing why and how these techniques result in asymptotically optimal systems is addressed in the Methods section for a specific choice of parameters, with a sketch of the proof provided in the SI appendix and proven more generally in (53).

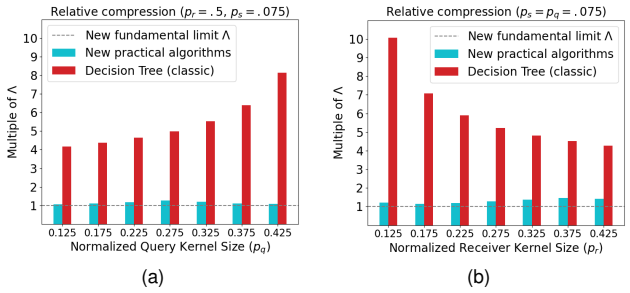


Fig. 3. Experimental results for two scenarios. (a) Results as a multiple of the new fundamental limit for the case Alice knows R, $p_r = 0.5, p_s = 0.075$ and $0.125 \leq p_q \leq 0.425$, demonstrating the significant gains in communication cost using practical semantic communication algorithms compared to purely classical approaches. (b) Results as a multiple of the new fundamental limit for the setting $p_r < 0.5$ and $p_s = p_q = 0.075$ in the case Alice does not know R, also comparing a classical approach with one leveraging logical semantics.

Empirical validation. In Figure 3, we present experimental results of practical algorithms for two contrasting situations. On the left, Alice and Bob have access to R, and Alice will help Bob prove a sentence Q, which is narrower than S. The sentence Q has an expected normalized kernel size of p_q, which we vary in the experiment. The red bars correspond to a coding technique based on an optimized representation (decision trees, explained in the subsection *Method of Generating Generalized Decision Tree Sentences* of the SI appendix) of either the sender or query logic sentences, whichever is cheapest after being passed through a good off-the-shelf lossless compressor (classical compression). The light-blue bars correspond to practical algorithms that we describe in subsections *Algorithms for targeted queries based on linear algebra* and *An algorithm for the case that Alice does not know what it is that Bob knows* of the Methods section, and which leverage the mathematics of Galois fields (58). Both sets

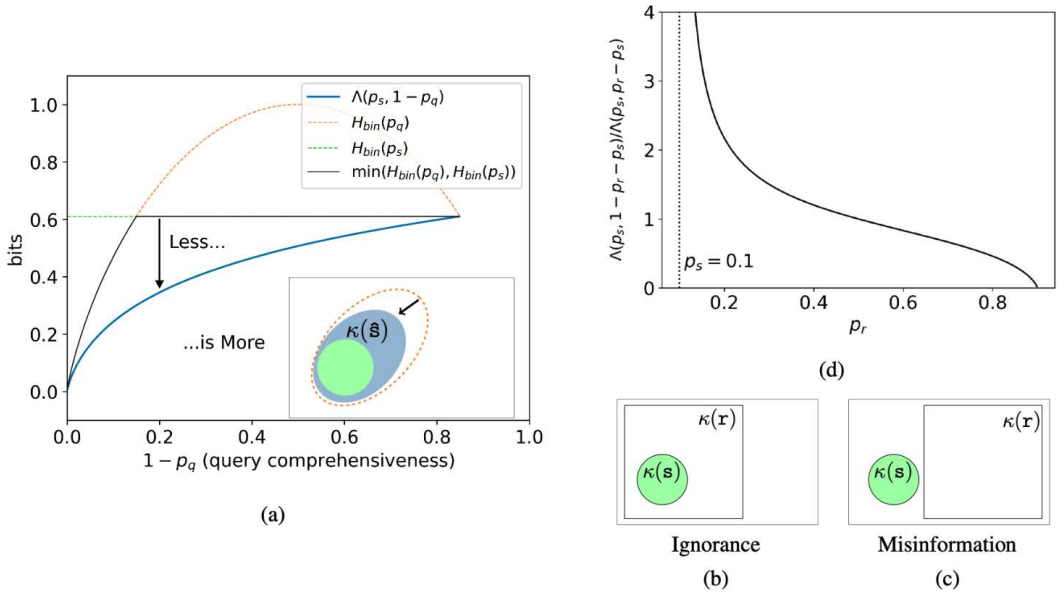


Fig. 4. (**a**) *Less is More.* In blue, the ultimate communication limit $\Lambda(\cdot,\cdot)$ for the case $p_r = 1$, as the query ranges from trivial ($p_q = 1$) to coinciding with the sender's information ($p_q = p_s = 0.15$). $\Lambda(\cdot,\cdot)$ is cheaper (Less...) than the two obvious strategies, yet the kernel size received by Bob is smaller than that of the query, showing Bob can prove even more things (is More...) than required. A similar picture will hold for any p_r. (**b**)-(**c**) *Kernel diagrams* that model Bob's state of ignorance versus incorrect information. (**d**) *Relative cost of incorrect information to ignorance.* We compare the ultimate limits for (b) (ignorance) and (c) (incorrect information); as Bob becomes more opinionated ($p_r \to p_s$, $p_s = 0.1$), the ratio goes to infinity.

of bars are plotted relative to the ultimate information-theoretic bound Λ, shown by the dashed line.

Our practical algorithms use a combination of novel ideas based on linear algebra over the Galois Field with two elements $\{0, 1\}$ together with the exploitation of classical coding techniques as described in the Methods section. To implement the competing classical approaches, we developed an optimized representation of logic sentences based on decision trees (59–61).

Of note is the significant savings possible by using techniques that take advantage of logic sentences and reasoning. In the data compression field, significant advances are often measured in improvements of a few percentage points whereas here we are illustrating gains in integer multiples.

No Need to Know. A striking consequence of Theorem 1 is that, in this model, the same communication-cost limit applies regardless of whether or not Alice knows what it is that Bob knows. This phenomenon is also present in lossless classical communication (62), but it is rarer in the case of lossy compression (63, 64). In our work, communication is inherently lossy, in that the logic sentence $\hat{\mathtt{S}}$ reproduced by Bob need not resemble Alice's $\mathtt{S}$ nor $\mathtt{Q}$.

The Less is More paradox. This is the scenario of a query $\mathtt{Q}$ that, while provably entailed by $\mathtt{S}$, is not equivalent to $\mathtt{S}$. Our result holds in the case Alice knows $\mathtt{R}$, for every choice of p_s, p_r and p_q such that $0 < p_s < p_q < p_r \leq 1$. The fundamental bound in this case is $\Lambda(p_s, p_r - p_q)$, which is in general more efficient than sending either $\mathtt{S}$ or $\mathtt{Q}$, which would respectively cost $p_r H_{\text{bin}}(p_s) = \Lambda(p_s, p_r - p_s)$ or $p_r H_{\text{bin}}(p_q) = \Lambda(p_q, p_r - p_q)$. In the upper part of Figure 4(a), we plot the information-theoretic limit in this scenario (in blue) and compare it against the bits that would be required to send either the query or the sender information, for the case $p_r = 1$. The fundamental limit is lower than both of these (Less...). Yet, when one examines the nature of the solution, one realizes that Bob obtains a sentence $\hat{\mathtt{S}}$ whose kernel is generally smaller than that of $\mathtt{Q}$, demonstrating that Bob is able to prove more than what he needed (...is More). The paradox is explained by noting that the way this is achieved is through deciding, ahead of time, on a terse collection of approximating kernels each of which can be used to solve the problem for multiple combinations of $\mathtt{S}$ and $\mathtt{Q}$, further reducing the communication cost; a small concrete example showing this phenomenon is given in a subsection *An example of a deterministic, optimal algorithm for a small problem with targeted queries* in the SI appendix. Of note, the same paradox can be seen as having potential implications for security – being as efficient as one can to allow Bob to prove $\mathtt{Q}$ using facts consistent with Alice's $\mathtt{S}$ results in revealing more than $\mathtt{Q}$. In another play on words, one may say that one needs to say more to say less.

The price of incorrect information. In a line of work not directly covered by Theorem 1, but exploiting the same ideas, we introduce a rudimentary model of incorrect information. Formally, we modify the setup as follows. The entailment assumption $\mathtt{S} \vdash \mathtt{R}$ is replaced by an *inconsistency* assumption:

$$\kappa(\mathtt{S}) \cap \kappa(\mathtt{R}) = \emptyset,$$

meaning that no model simultaneously satisfies both Alice's and Bob's sentences. We assume Alice knows $\mathtt{R}$, and that both parties agree Alice possesses the correct facts. The system succeeds if Bob obtains $\hat{\mathtt{S}}$ logically equivalent to $\mathtt{S}$, i.e., $\kappa(\hat{\mathtt{S}}) = \kappa(\mathtt{S})$. The figure of merit remains $\min_{f,g} E[\mathbf{len}(f(\mathtt{S}, \mathtt{R}))]/|\mathcal{M}|$, with the probabilistic model unchanged except that the disjointness constraint requires $p_s + p_r \leq 1$ (Figure 4(c)). In contrast, for the setting studied elsewhere in this article (Figure 4(b)), Bob may

be said to be in a state of ignorance, where $\kappa(\mathtt{S}) \subseteq \kappa(\mathtt{R})$, rather than having incorrect information.

Clearly this model is an unrealistic cartoon of the complex problem of incorrect information; we introduce it here merely to explore the type of conclusions one can derive from it, fully aware of its limitations; see (33) for an early occurrence of this general type of problem. The limit $\Lambda(p_s, 1-p_r-p_s)$ can be understood by observing that disjointness implies $\kappa(\mathtt{S}) \subseteq \kappa(\neg\mathtt{R})$; that is, the negation of Bob's incorrect sentence, with normalized kernel size $1-p_r$, serves as effective background knowledge logically consistent with Alice's $\mathtt{S}$. Applying the same reasoning as in Theorem 1 with this substitution yields the result. This is admittedly a crude bound—the negation of Bob's beliefs is far from ideal as background knowledge—but it suffices to illustrate how the cost of correction escalates dramatically. For ignorance, as previously discussed, the limit is $\Lambda(p_s, p_r - p_s)$. We plot their ratio as a function of p_r (Figure 4(d)) showing that as Bob becomes more opinionated—meaning p_r decreases toward p_s, so that his (incorrect) sentence imposes more restrictions on the world—the relative cost of correcting incorrect information versus ignorance grows to infinity, a result that appears to agree with common intuition.

Methods

In this section we first describe the development of the practical algorithms used in our experiments. We then describe the experimental setup, and provide an expanded discussion on the experimental results. These experimental methods can attain the ultimate limits discovered in this work only under very specific circumstances, leaving open the task of finding algorithms that approach such limits more generally.

Overview of algorithms. We have devised two practical algorithms that are components of our proposed extended communication system, which are respectively aimed at the following scenarios:

- The "targeted query" scenario where both Alice and Bob know $\mathtt{R}$ and Alice must send the minimum amount of bits to Bob to allow him to prove a query $\mathtt{Q}$ that she can prove; and
- The scenario where Bob knows $\mathtt{R}$ but Alice does not know it; in this case the goal is for Bob to be able to prove all that Alice can while making the total exchange of bits between them, potentially over multiple turns and in any direction, as small as possible.

The algorithms for these two scenarios are based on the mathematics of linear algebra over the Galois Field (58) with two elements $\{0,1\}$, denoted by $GF(2)$, and are shown to be optimal in specific settings. We then present experimental results comparing the performance of an ensemble of methods, including those developed herein, and contrast them with the Λ bound in Theorem 1 as well as optimized methods that treat the logic expressions as strings in the classical information-theoretic sense.

All of our algorithms target the efficient representation and decoding of kernels directly, and thus assume the prior existence of functions that map logic sentences to kernels and vice-versa. Note that while kernels are, strictly speaking, subsets of $\mathcal{M}$, it is possible to represent them as binary vectors in $\{0,1\}^{|\mathcal{M}|}$. Thus all of our algorithms will be phrased as a means to find encoders and decoders with various desirable properties for such binary vectors.

For two discrete random variables A, B with joint probability mass function $P(A,B)$, we respectively define entropy and conditional entropy as

$$H(A) := E\left[\log \frac{1}{P(A)}\right], H(A \mid B) := E\left[\log \frac{1}{P(A \mid B)}\right].$$

Furthermore, we define Shannon's binary entropy by

$$H_{\text{bin}}(p) := \begin{cases} -p\log p - (1-p)\log(1-p), & p \in (0,1), \\ 0, & \text{otherwise.} \end{cases}$$

Recall that $\mathcal{M}$ is a finite set of models, and that the kernel of a logic sentence, $\kappa(\cdot)$, is the set of models that make that sentence true. We assume an enumeration of such models:

$$\mathcal{M} = \{\mu_1, \ldots, \mu_{|\mathcal{M}|}\}.$$

Conditions for the validity of the lower bound of Theorem 1. In the design of our experiments, we seek to demonstrate the extent to which our practical methods can achieve the lower bound of Theorem 1. Here, we rigorously complete the statement of this theorem by specifying the conditions needed for the validity of this lower bound, which will inform how we design the experiments. First, recall that the upper bound of Theorem 1 holds in significant generality, assuming only that

- $\mathtt{S} \vdash \mathtt{Q}$, $\mathtt{Q} \vdash \mathtt{R}$ (and therefore $\mathtt{S} \vdash \mathtt{R}$).

Define, temporarily, the random vector $\vec{\kappa}(\mathtt{S})$ to be an indicator vector for the kernel of $\mathtt{S}$, governed by the rule $\vec{\kappa}(\mathtt{S})_i = 1$ if $\mu_i \in \kappa(\mathtt{S})$ and 0 otherwise, and similarly define the same for $\mathtt{Q}$ and $\mathtt{R}$. The lower bound augments the above assumptions on the upper bound with the additional assumptions:

- The random tuples $\{(\vec{\kappa}(\mathtt{S}), \vec{\kappa}(\mathtt{Q}), \vec{\kappa}(\mathtt{R}))_i\}$ are i.i.d.;
- $\mathtt{R}$ and $\kappa(\mathtt{S})$ are conditionally independent given $\kappa(\mathtt{R})$.

Algorithms for targeted queries based on linear algebra. Our treatment of the problem of targeted queries allows for a sentence $\mathtt{R}$ to be optionally shared by both Alice and Bob. We present the algorithm in the setting where there is no such $\mathtt{R}$ for didactic reasons; extension to the case where such a sentence is available to both Alice and Bob is relatively straightforward. The algorithm starts by Alice computing the kernels $\kappa(\mathtt{S}) \subseteq \mathcal{M}$ and $\kappa(\mathtt{Q}) \subseteq \mathcal{M}$; Alice accomplishes her goal if she is able to send to Bob a subset of $\mathcal{M}$ that includes $\kappa(\mathtt{S})$ and is included in $\kappa(\mathtt{Q})$. In this algorithm, we let $X_i = 1$ if $\mu_i \in \kappa(\mathtt{S})$, $X_i = 0$ if $\mu_i \in \kappa(\mathtt{Q})^c$ and $X_i = 2$ otherwise; this then defines a vector $X^{|\mathcal{M}|} \in \{0,1,2\}^{|\mathcal{M}|}$.

Notice that if we were able to find a vector $\hat{X}^{|\mathcal{M}|} \in \{0,1\}^{|\mathcal{M}|}$ which agrees with $X^{|\mathcal{M}|}$ on all the positions $\{i : X_i \in \{0,1\}\}$, then Alice would accomplish her goal, since such a vector can be seen as denoting the kernel of a sentence that Bob will be recovering from Alice's transmission; note that such a vector would in effect be a partition of the sets $\{i : X_i = 0\}$ and $\{i : X_i = 1\}$.

Furthermore, to accomplish her goal economically, there must be a way to transmit this vector efficiently to Bob, where efficiency is quantified as the expected number of bits transmitted.

We now present a general algorithm for finding partitions that can be represented efficiently by exploiting linear algebra over $GF(2)$. This algorithm can be applied even when we do not have any statistical model of the underlying sets for which we are creating a partition. In the special case where the entries of $X^{|\mathcal{M}|}$ are drawn i.i.d. from $\{0,1,2\}$ using probability masses $(p_0, p_1, 1-p_0-p_1)$, we will show that this algorithm is asymptotically optimal if $p_0 = p_1$, when competing against the class of all possible algorithms, whether based on linear concepts or not. Note that there is no special practical significance to this case; the existence of algorithms achieving the lower bound of Theorem 1 in complete generality is proven in (53), with a proof sketch given in the SI appendix.

The practical algorithm assumes that Alice and Bob share a means of creating identical pseudorandom bits independently; this can be easily achieved by initially agreeing to a pseudorandom number generator and a seed. For ease of analysis, we assume that such pseudorandom bits are i.i.d. bits uniformly distributed over $\{0,1\}$.

Taking $|\mathcal{M}|$ of these random bits we can construct a $1 \times |\mathcal{M}|$ row vector which we shall denote by G_1. Taking another $|\mathcal{M}|$ of them, we can append a second row to G_1, leading to a $2 \times |\mathcal{M}|$ matrix that we denote by G_2. By repeating this procedure an arbitrary, but finite, number of times, we can assume that Alice and Bob share G_r for any r. In practice, the number of rows r we will be using, although variable, will be slightly larger than $|\mathcal{M}|(p_0+p_1)$. Define

$$\Psi := \{i : X_i \in \{0,1\}\}.$$

For any given vector $x \in \{0,1\}^{|\mathcal{M}|}$ and set of indices $\xi \subseteq \{0,1,\ldots,|\mathcal{M}|-1\}$, let $[x]_\xi$ denote the $|\xi|$-long vector obtained by extracting from x only the indices given by ξ. The algorithm starts by finding the smallest positive integer J such that the equation

$$[M \cdot G_J]_\Psi = [X^{|\mathcal{M}|}]_\Psi, \qquad \textbf{[2]}$$

can be solved for some $1 \times J$ vector M with entries in $GF(2)$; here, the matrix multiplication is interpreted to be using $GF(2)$ arithmetic. Intuitively, such a solution can be found because the dimension of the row space of G_j increases as j grows with high probability; thus this dimension will eventually become $|\Psi|$, rendering the equation always solvable; this observation will be shortly used formally.

Once Alice finds such a J, she first sends it and then sends M. We could represent J in base 2 and transmit such bits; however, this presumes that both sender and receiver know, ahead of time, how many bits are needed to represent J. A more general solution can be found from the literature of *universal codes for integers*; one such example is a code due to Elias (65) which represents an arbitrary positive integer j with

$$\lfloor \log_2 j \rfloor + 2\lfloor \log_2(1 + \lfloor \log_2(j) \rfloor) \rfloor + 1 \leq \log_2 j + 2\log_2(1 + \log_2 j) + 1 \qquad \textbf{[3]}$$

bits, where the upper bound is valid for $j \geq 1$. Denoting this code as $\text{elias}_\delta(j)$, and the number of bits that such a code assigns to j by $\mathbf{len}(\text{elias}_\delta(j))$, the sender then sends the integer J to the receiver using

$$\mathbf{len}(\text{elias}_\delta(J))$$

bits, followed by the J bits in the message M. The receiver then decodes J and computes $M \cdot G_J$ to retrieve the partition.

We now analyze the expected performance of this algorithm. We first observe that

$$\begin{aligned} & E\big[J + \mathbf{len}(\text{elias}_\delta(J))\big] - 1 \\ & \overset{(a)}{\leq} E\big[J + \log_2 J \quad + 2\log_2(1 + \log_2 J)\big] \\ & \overset{(b)}{=} E\big[E\big[J + \log_2 J + 2\log_2(1 + \log_2 J) \mid \Psi\big]\big] \\ & \overset{(c)}{\leq} E\big[E[J \mid \Psi] + \log_2 E[J \mid \Psi] + \\ & 2\log_2(1 + \log_2 E[J \mid \Psi])\big] \end{aligned} \qquad \textbf{[4]}$$

where (a) follows from Eq. (3), (b) follows from the law of total expectation, and (c) follows from the concavity of the logarithm. We next upper bound $E[J \mid \Psi]$. To obtain an upper bound for J, we note that a sufficient condition to be able to solve Eq. (2) is that $[G_J]_\Psi$ has full-row rank, that is, the dimension of the space spanned by the rows of $[G_J]_\Psi$ is exactly $|\Psi|$.

Let W_i be the smallest integer such that the row subspace spanned by $[G_{W_i}]_\Psi$ has dimension i. We can trivially write

$$W_{|\Psi|} = W_1 + \sum_{i=1}^{|\Psi|-1} W_{i+1} - W_i.$$

The probability that a vector drawn uniformly from $GF(2)^{|\Psi|}$ is nonzero, and therefore spans a space of dimension 1, is given by $1 - 2^{-|\Psi|}$; hence,

$$E[W_1 \mid \Psi] = \frac{1}{1 - 2^{-|\Psi|}}.$$

We now analyze the expected difference $E[W_{i+1} - W_i \mid \Psi]$. Note that, by definition, $[G_{W_i}]_\Psi$ spans a subspace of dimension i, which must consist of exactly 2^i elements. If one chooses an element from $GF(2)^{|\Psi|}$ uniformly at random, the probability that it lies within the subspace spanned by $[G_{W_i}]_\Psi$ is $2^{i-|\Psi|}$. As a consequence, the probability that the subspace spanned by the rows of $[G_{W_i}]_\Psi$ with such a random vector having dimension $i+1$ is $1 - 2^{i-|\Psi|}$, and thus

$$E[W_{i+1} - W_i \mid \Psi] = \frac{1}{1 - 2^{i-|\Psi|}}.$$

Observing that $J \leq W_{|\Psi|}$, we have

$$E[J \mid \Psi] \leq \sum_{i=0}^{|\Psi|-1} \frac{1}{1 - 2^{i-|\Psi|}} \leq |\Psi| + 2.$$

Continuing from the inequality in Eq. (4), and normalizing by $|\mathcal{M}|$, we obtain the following upper bound on the bit

cost:

$$E\big[|\Psi| + \log_2(|\Psi|+2) + 2\log_2\left(1+\log_2(|\Psi|+2)\right)+3\big]/|\mathcal{M}|.$$

To understand the quality of this algorithm, we assume that the entries of $X^{|\mathcal{M}|}$ are drawn i.i.d. from $\{0,1,2\}$ using probability masses $(p_0, p_1, 1-p_0-p_1)$ and therefore $E[|\Psi|] = |\mathcal{M}|(p_0+p_1)$, which results in the following upper bound on average performance:

$$p_0+p_1+\frac{\log_2\left(|\mathcal{M}|(p_0+p_1)+2\right)}{|\mathcal{M}|} + 2\frac{\log_2\left(1+\log_2\left(|\mathcal{M}|(p_0+p_1)+2\right)\right)}{|\mathcal{M}|}+\frac{3}{|\mathcal{M}|}. \quad [5]$$

A special case of mathematical, rather than practical, interest arises when $p_0 = p_1$. The achievable limit for this setting, as unveiled in our work, is given by

$$\Lambda(p_0,p_1) = \Lambda(p_0,p_0) = 2p_0,$$

which is asymptotically, as $|\mathcal{M}| \to \infty$, the same performance as in Eq. (5) and thus our algorithm is asymptotically optimal. If $p_0 \neq p_1$, then $\Lambda(p_0,p_1) < p_0+p_1$ and thus our algorithm is not asymptotically optimal. For illustration purposes, the reader will find an example of an optimal code for a small setting in the SI appendix. In the extended article (53), we present an asymptotically optimal general construction based on random coding arguments.

An algorithm for the case that Alice does not know what it is that Bob knows. As a parallel to the previous subsection, we start by assuming that Alice computes the kernel $\kappa(\mathtt{S}) \subseteq \mathcal{M}$ and that Bob computes the kernel $\kappa(\mathtt{R}) \subseteq \mathcal{M}$. Alice will accomplish her goal if she is able to transmit $\kappa(\mathtt{S})$ efficiently, somehow leveraging the fact that she knows that Bob knows an $\mathtt{R}$ with the property that $\kappa(\mathtt{S}) \subseteq \kappa(\mathtt{R})$, but not knowing anything further about $\mathtt{R}$ or $\kappa(\mathtt{R})$ at the outset of the communication.

In this algorithm, we define the vectors $X^{|\mathcal{M}|}, Y^{|\mathcal{M}|}$ through the following: $X_i = 1$ if $\mu_i \in \kappa(\mathtt{S})$ and $X_i = 0$ otherwise; similarly, $Y_i = 1$ if $\mu_i \in \kappa(\mathtt{R})$ and $Y_i = 0$ otherwise; note that, by assumption, $\{i : X_i = 1\} \subseteq \{i : Y_i = 1\}$. We assume that Alice knows $X^{|\mathcal{M}|}$ but not $Y^{|\mathcal{M}|}$ (other than the condition above), and that Bob knows $Y^{|\mathcal{M}|}$. The goal is to efficiently transmit $X^{|\mathcal{M}|}$ to Bob.

Let $\Delta > 0$ be an integer that is a design parameter. As in the previous algorithm using linear methods, we assume that Alice and Bob share a means for generating the same pseudorandom bits. In our analysis we assume that Alice and Bob share a matrix G with dimension $(|\mathcal{M}|+\Delta)\times|\mathcal{M}|$ and entries drawn uniformly and independently at random from $GF(2)$. We use the notation G_r to denote the first r rows of the matrix G.

Let $\Psi = \{i : Y_i = 1\}$. The method starts with Bob transmitting to Alice the integer $|\Psi|$. At this point both Alice and Bob will keep only the first $|\Psi|+\Delta$ rows of G, denoted by $G_{|\Psi|+\Delta}$. Alice sends to Bob the bits resulting from the matrix-vector multiplication $G_{|\Psi|+\Delta}X^{|\mathcal{M}|}$, where $X^{|\mathcal{M}|}$ is interpreted as a column vector. Let $[G_{|\Psi|+\Delta}]_\Psi$ denote the matrix obtained by extracting from $G_{|\Psi|+\Delta}$ the columns implied by the indices Ψ; note that this matrix is computable by Bob but not by Alice. Bob then attempts to solve the equation

$$[G_{|\Psi|+\Delta}]_\Psi z = G_{|\Psi|+\Delta}X^{|\mathcal{M}|} \quad [6]$$

for a unique z. If such a unique z exists, Bob can retrieve $X^{|\mathcal{M}|}$ by making use of the fact that the only entries of $X^{|\mathcal{M}|}$ which could possibly be equal to 1 must have an index contained in Ψ and the fact that

$$[X^{|\mathcal{M}|}]_\Psi = z.$$

Therefore, $X^{|\mathcal{M}|}$ can be recovered from z by lifting the latter using the indices Ψ.

By construction, Eq. (6) has at least one solution. If the $|\Psi|$ columns of $[G_{|\Psi|+\Delta}]_\Psi$ are linearly independent, then that solution must be unique. The probability that these columns are linearly independent is given by $\prod_{i=0}^{|\Psi|-1}(1-2^{i-|\Psi|-\Delta})$ and can be lower bounded in the following manner:

$$\begin{aligned}
\prod_{i=0}^{|\Psi|-1}\left(1-2^{i-|\Psi|-\Delta}\right) &= \exp\Big(\sum_{i=0}^{|\Psi|-1}\log\left(1-2^{i-|\Psi|-\Delta}\right)\Big)\\
&\geq \exp\Big(\sum_{i=0}^{|\Psi|-1}-\tfrac{2^{i-|\Psi|-\Delta}}{1-2^{i-|\Psi|-\Delta}}\Big) = \exp\Big(\sum_{i=1}^{|\Psi|}-\tfrac{2^{-i-\Delta}}{1-2^{-i-\Delta}}\Big)\\
&\geq \exp\Big(\sum_{i=1}^{|\Psi|}-2^{-i-\Delta+1}\Big) = \exp\Big(-2^{-\Delta+1}\sum_{i=1}^{|\Psi|}2^{-i}\Big)\\
&\geq \exp\left(-2^{-\Delta+1}\right).
\end{aligned}$$

Bob can signal to Alice success or failure in finding a unique z with a single bit, and in the case of failure, Alice can simply send $X^{|\mathcal{M}|}$ verbatim by sending $|\mathcal{M}|$ bits.

The expected number of bits transmitted in either direction is then upper bounded as follows:

$$\begin{aligned}
\text{cost} &= 1+(\text{probability of failure})\,|\mathcal{M}| + E\left[|\Psi|+\Delta\right]\\
&\quad + E\left[\mathbf{len}(\text{elias}_\delta(|\Psi|))\right]\\
&\leq 1+\left(1-\exp\left(-2^{-\Delta+1}\right)\right)|\mathcal{M}| + E\left[|\Psi|\right]+\Delta\\
&\quad + E\left[\mathbf{len}(\text{elias}_\delta(|\Psi|))\right]\\
&\leq \left(1-\exp\left(-2^{-\Delta+1}\right)\right)|\mathcal{M}| + E\left[|\Psi|\right]+\Delta\\
&\quad + \log_2 E\left[|\Psi|\right] + 2\log_2\left(1+\log_2 E[|\Psi|]\right)+2.
\end{aligned}$$

We now make the choice $\Delta = \lceil\log_2|\mathcal{M}|\rceil$; using the inequality $1+x \leq \exp(x)$ with $x = -2^{-\Delta+1}$ yields

$$1-\exp\left(-2^{-\Delta+1}\right) \leq 2^{-\Delta+1} \leq \frac{2}{|\mathcal{M}|}.$$

Normalizing this cost upper bound by $|\mathcal{M}|$ and defining $q_1 := E\left[|\Psi|\right]/|\mathcal{M}|$, the normalized upper bound may be simplified to

$$q_1 + O\left(\frac{\log_2|\mathcal{M}|}{|\mathcal{M}|}\right), \quad [7]$$

as $|\mathcal{M}| \to \infty$. To understand the quality of the above bound, assume temporarily that $Y^{|\mathcal{M}|}$ has entries drawn

i.i.d. from $\{0,1\}$ using probability masses $(1-q_1, q_1)$, and that each X_i is drawn using the conditional distribution

$$\begin{aligned} P(X_i = 1 | Y_i = 1) &= p_1/q_1, \\ P(X_i = 0 | Y_i = 0) &= 1. \end{aligned}$$

Under these assumptions for $X^{|\mathcal{M}|}, Y^{|\mathcal{M}|}$, a lower bound is given by $|\mathcal{M}|^{-1} H(X^{|\mathcal{M}|} | Y^{|\mathcal{M}|}) = q_1 H_{\text{bin}}(p_1/q_1)$. Comparing this lower bound to Eq. (7) we see that, if $q_1 = 2p_1$ and $|\mathcal{M}|$ is sufficiently large, then the linear coding algorithm can arbitrarily approach the lower bound; this observation is made for mathematical completeness as there is no special practical importance to this setting. For other choices of q_1, p_1, nonetheless the algorithm is not optimal. As before, the existence of asymptotically optimal algorithms for all choices of parameters is proved in the extended article (53), with a proof sketch given in the SI appendix.

Experimental setup. The goal of this section is to illustrate the possible gains that one may expect from a practical semantic communication system compared to a classical one, and to also show the existing gap of such a semantic communication system with respect to the ultimate bound given by Λ.

A first problem is the fact that it is possible to portray classical communication systems as nearly arbitrarily inefficient when compared to semantic ones. The reason is due to the multiple different ways in which sentences can express the same underlying semantic content; classical compression systems must be faithful to the original sentence itself whereas semantic systems, as treated in this article, can take advantage of the intended meaning of the symbols within the sentence. To illustrate this problem, given $H(\kappa(\mathtt{S}) \,|\, \mathtt{S}) = 0$, we note that

$$H(\mathtt{S}) = H(\mathtt{S} \,|\, \kappa(\mathtt{S})) + H(\kappa(\mathtt{S})) \geq H(\kappa(\mathtt{S})). \qquad [8]$$

The gap $H(\mathtt{S} \,|\, \kappa(\mathtt{S}))$ can be very large; consider, for example, propositional logic in two variables $\mathtt{Z}_1, \mathtt{Z}_2$ and observe that

$$\begin{gathered} \mathtt{Z}_3 \to (\mathtt{Z}_1 \wedge \mathtt{Z}_2) \\ (\mathtt{Z}_1 \wedge \mathtt{Z}_2) \vee \neg \mathtt{Z}_3 \\ \neg(\neg(\mathtt{Z}_1 \wedge \mathtt{Z}_2) \wedge \mathtt{Z}_3) \\ \neg((\neg \mathtt{Z}_1 \vee \neg \mathtt{Z}_2) \wedge \mathtt{Z}_3) \\ \neg((\neg \mathtt{Z}_1 \wedge \mathtt{Z}_3) \vee (\neg \mathtt{Z}_2 \wedge \mathtt{Z}_3)) \\ \neg((\neg \mathtt{Z}_1 \wedge \mathtt{Z}_3) \vee (\neg \mathtt{Z}_2 \wedge \mathtt{Z}_3)) \wedge (\mathtt{Z}_1 \vee \neg \mathtt{Z}_1) \end{gathered}$$

are all logically equivalent sentences, yet appear different when interpreted as simple strings. One can expect semantic communication systems to take advantage of the observation in Eq. (8) and achieve a performance that is roughly $H(\kappa(\mathtt{S}))$, but one must exercise caution and not overstate the benefit; for example, recent estimates (66) assert that in an average English sentence, about half of the bits can be attributed to establishing meaning (so to $H(\kappa(\mathtt{S}))$), and the other half to the specific choice of wording (so, roughly, to $H(\mathtt{S} \,|\, \kappa(\mathtt{S}))$). Our paper's theoretical results not only leverage the phenomenon in Eq. (8), but go beyond and exploit more delicate findings on how targeted queries may admit unusually efficient representations, or how, in some occasions, communication is surprisingly just as efficient when Alice does not know Bob's sentence $\mathtt{R}$ as when she does know $\mathtt{R}$.

Our experimental evaluation methodology follows the philosophy of always choosing a stronger "classical" (non-semantic) baseline to compare against whenever a choice is on the table, even if these baselines start to become more semantic in nature. In particular:

- On purpose, we forego the type of advantage that stems from the observation in Eq. (8) even though we can legitimately claim it.
- Logic sentences will be optimized so that they can be represented more compactly by "classical" compressions systems.

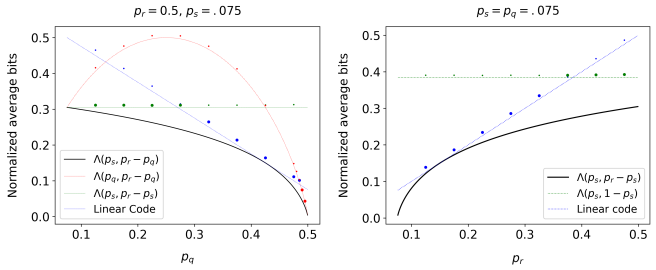
Scenarios. We demonstrate experimental results for the following two contrasting scenarios:

- **Targeted query with a shared sentence.** After the communication, Bob will be able to prove $\mathtt{Q}$, which satisfies $\mathtt{S} \vdash \mathtt{Q}$. Both Alice and Bob know $\mathtt{R}$; only Alice knows $\mathtt{S}$.
- **Alice does not know what it is that Bob knows.** After the communication, Bob will be able to prove all that Alice can. Only Bob knows $\mathtt{R}$, and only Alice knows $\mathtt{S}$.

Test case generation. For the purposes of our experiments, we first choose the underlying kernels of $\mathtt{S}, \mathtt{Q}, \mathtt{R}$, and then generate the corresponding sentences; we stress that this is done for convenience and in no way reflects limitations on Theorem 1 which is stated for general distributions, particularly for the upper bound. We generated random kernels meeting all three constraints above on sets of 10 variables to test our methods in practice. For each of the kernels, we produced a generalized decision tree with the given kernel as its set of satisfying truth-value assignments. We then post-processed each sentence by writing it using postfix notation, compressing variable names, and removing spaces. For the targeted query scenario, we have chosen $p_r = 0.5$, $p_s = 0.075$ and p_q by sampling the range $[p_s, p_r)$ uniformly. For the scenario where Alice does not know what it is that Bob knows, we set $p_s = p_q = 0.075$ and let p_r range in $(p_s, 0.5]$. We generated 1000 test cases for each choice of (p_s, p_q, p_r) considered. For specifics, we refer the reader to the *Additional details on experimental methodology* section of the SI appendix.

Experimental results

Semantic communication systems. In a first set of experiments, we aimed to demonstrate the performance of our practical semantic communication techniques against the theoretical bounds. We have implemented the linear codes introduced earlier in this Methods section, which are used to address the two respective scenarios. Due to the fact that these linear codes are not always optimal, we augmented these systems with "näive" semantic communication strategies based on efficient lossless transmission of kernels. For certain parameter ranges these are better than the linear codes, and thus in such cases we can use them instead. In the case of targeted queries, the task of allowing Bob to prove $\mathtt{Q}$ can alternatively be accomplished by sending $\kappa(\mathtt{Q})$ or $\kappa(\mathtt{S})$ to him, whichever is cheapest. The results, averaged across all 1000 test cases for each



(a) Semantic communication systems for targeted queries with a shared sentence with $p_r = 0.5, 0.075 < p_q \leq 0.5$.

(b) Semantic communication systems for when Alice does not know what it is that Bob knows with $p_s = p_q = 0.075$.

Fig. 5. Comparison of practical semantic communication methods against the new $\Lambda(p_s, p_r - p_q)$ limit derived in this work for two scenarios.

choice of parameters, can be found in Figure 5(a) where in blue we illustrate the performance of linear codes, and in green and red we respectively illustrate the performance of sending $\kappa(\mathtt{S})$ and $\kappa(\mathtt{Q})$ losslessly using such efficient techniques. These methods rely on the idea of enumerating all possible kernels that can be transmitted and then computing a code directly from the integer representing the position in which such a kernel appears in the list, rather than the kernel itself. For an overview of these codes based on enumerative techniques, we refer the reader to (67).

In each of these three practical systems, we include a curve of the same color that is a lower bound on performance for the specific technique. The gap between the dots and the lower bound curves is primarily due to the added cost of the Elias code which becomes negligible as the length of the code increases. The bolder dots are those that are closest to the fundamental limit derived in our article, which is the bold, lowest curve shown in the figure. Similarly, in the case that Alice does not know what it is that Bob knows, the task of allowing Bob to prove $\mathtt{S}$ can alternatively be accomplished by sending $\kappa(\mathtt{S})$ to him, also using enumerative source coding (67). The results of this experiment can be found in Figure 5(b). Our practical codes can be quite efficient in certain scenarios, but additional work is needed to develop practical codes that come closer to the fundamental limits for wider ranges of p_q and p_r; refer to Figure 5(a) and Figure 5(b), respectively.

Classical communication systems. In a second set of experiments, we aimed to contrast practical classical systems with semantic ones. In these experiments, we reused the best results from the practical methods above and compared them with classical compression systems.

We define classical compression systems as ones that may only use the sentence as presented to the communication system, and not perform operations on it with awareness of its underlying semantic content. These are examples of such systems:

- **Targeted query with a shared sentence.** Using the Decision Tree representation of a sentence, employ a standard compression algorithm (in particular, gzip, lzma, bzip2) to compress $\mathtt{S}$ and $\mathtt{Q}$; choose the best representation and send that one.
- **Alice does not know what it is that Bob knows.** The same as above, but considering only $\mathtt{S}$ as in this scenario, $\mathtt{Q} = \mathtt{S}$.

We note that the classical system is forced to solve a fundamentally stricter problem, lossless reconstruction of the full sentence, because it lacks the capacity to make logical deductions. A system that has this capability can tolerate a lossy reconstruction as long as it is possible to deduce $\mathtt{Q}$ from $\mathtt{R}$ together with the information transmitted. We provided an additional advantage to the type of classical communication systems used above: we compressed all 1000 samples simultaneously, which allowed the compression algorithms to leverage patterns that only emerge when more data are available. In contrast, the semantic communication systems compress just one instance at a time. We reiterate that by using Decision Trees, we have already used semantic concepts to benefit the classical baseline. The results corresponding to these experiments can be found in Figures 3(a) and 3(b) above, as described in the main body of this article. These plots normalize the performance of the practical semantic or classical systems against the corresponding fundamental limit $\Lambda(p_s, p_r - p_q)$ and $\Lambda(p_s, p_r - p_s)$, respectively. We acknowledge that in the realm of classical communication systems we should consider Slepian-Wolf compression algorithms (62) as a baseline. This is particularly difficult to do as we are not aware of practical instances of such algorithms that would be applicable to the complex statistical distributions present in $\mathtt{S}$ and $\mathtt{R}$; we leave this comparison for future research.

Implications and next steps. We have provided evidence that new communication system designs incorporating deductive inference mechanisms can be effectively analyzed and that the potential efficiency gains can be significant. To capitalize on this potential, nonetheless, additional progress is required.

First, our deductive inference-centered proposal can, in principle, be combined with other proposals (34, 35, 37, 38) that similarly rely on Shannon-theoretic principles, potentially unveiling even more communication efficiencies. Second, the communication model we introduced can be enhanced in many directions, including situations where Alice and Bob's roles are more symmetric (they each know something useful to a joint goal (68)), adversarial (one may be trying to convince the other about something that is not true), consultant (Bob is seeking help for a specific matter), multiparty (one instructor, many students), and machine-oriented (Alice and Bob are AI agents). Sharper results that forego the expected cost assumption and instead hold for individual logic sentences are possible; we anticipate that techniques such as the ones developed in (69) will be relevant in establishing these results.

We close with more speculative directions, each pointing to ways in which the interplay between logic and information theory may deepen. Just as standard information theory has provided mathematical foundations for machine learning (70), semantic information theory may provide analogous foundations for AI systems that incorporate logical reasoning. In (53) we show that our results hold for Propositional Logic and First-Order Logic over finite models. We anticipate that our results will be extended beyond finite model theory, and to richer logics beyond First-Order Logic, for example

to First-Order Logic with counting (71), Second-Order Logic (72, 73), or to logic capturing uncertainty (74–76). Given the close connections between Kolmogorov complexity and Shannon's entropy (77), we anticipate connections between semantic information theory and computer programs. Each of these directions reinforces the broader thesis that reasoning and information are deeply intertwined, and that formalizing their interaction can yield both theoretical and practical dividends.

ACKNOWLEDGMENTS. The authors acknowledge helpful conversations with the following individuals: Phokion Kolaitis, Jason Rute, Madhu Sudan, Kush Varshney, and Mark Wegman. Work of W. Szpankowski was partially supported by the NSF Center for Science of Information (CSoI) Grant CCF-0939370, and also by NSF Grants CCF-2006440, and CCF-2211423.

1. R Feynman, R Leighton, M Sands, E Hafner, *The Feynman Lectures on Physics; Vol. I.* (Addison-Wesley), (1965).
2. L Brillouin, Information theory and its applications to fundamental problems in physics. *Nature* **183**, 501–502 (1959).
3. RF Nalewajski, RG Parr, Information theory, atoms in molecules, and molecular similarity. *Proc. Natl. Acad. Sci.* **97**, 8879–8882 (2000).
4. D McLachlan, Crystal structure and information theory. *Proc. Natl. Acad. Sci.* **44**, 948–956 (1958).
5. G Hummer, S Garde, AE García, A Pohorille, LR Pratt, An information theory model of hydrophobic interactions. *Proc. Natl. Acad. Sci.* **93**, 8951–8955 (1996).
6. A Golan, J Harte, Information theory: A foundation for complexity science. *Proc. Natl. Acad. Sci.* **119**, e2119089119 (2022).
7. TD Ladd, et al., Quantum computers. *Nature* **464**, 45–53 (2010).
8. L Masanes, MP Müller, R Augusiak, D Pérez-García, Existence of an information unit as a postulate of quantum theory. *Proc. Natl. Acad. Sci.* **110**, 16373–16377 (2013).
9. JS Rowlinson, Probability, information and entropy. *Nature* **225**, 1196–1198 (1970).
10. T Raz, RD Levine, The essence of phase transitions in condensed matter by an information theoretic approach. *Proc. Natl. Acad. Sci.* **120**, e2310281120 (2023).
11. JA Wheeler, Information, physics, quantum: The search for links in *Proceedings III International Symposium on Foundations of Quantum Mechanics*, ed. WJ Archibald. pp. 354–358 (1989).
12. CRS Banerji, et al., Cellular network entropy as the energy potential in waddington's differentiation landscape. *Sci. Reports* **3**, 3039 (2013).
13. T Jetka, K Nienałtowski, S Filippi, MPH Stumpf, M Komorowski, An information-theoretic framework for deciphering pleiotropic and noisy biochemical signaling. *Nat. Commun.* **9**, 4591 (2018).
14. Y Tang, et al., Quantifying information accumulation encoded in the dynamics of biochemical signaling. *Nat. Commun.* **12**, 1272 (2021).
15. DB Brückner, G Tkačik, Information content and optimization of self-organized developmental systems. *Proc. Natl. Acad. Sci.* **121**, e2322326121 (2024).
16. O Martínez, MH Reyes-Valdés, Defining diversity, specialization, and gene specificity in transcriptomes through information theory. *Proc. Natl. Acad. Sci.* **105**, 9709–9714 (2008).
17. G Jenkinson, E Pujadas, J Goutsias, AP Feinberg, Potential energy landscapes identify the information-theoretic nature of the epigenome. *Nat. Genet.* **49**, 719–729 (2017).
18. J Henderson, Y Nagano, M Milighetti, A Tiffeau-Mayer, Limits on inferring t cell specificity from partial information. *Proc. Natl. Acad. Sci.* **121**, e2408696121 (2024).
19. XL Jiang, E Martinez-Ledesma, F Morcos, Revealing protein networks and gene-drug connectivity in cancer from direct information. *Sci. Reports* **7**, 3739 (2017).
20. RC deCharms, Information coding in the cortex by independent or coordinated populations. *Proc. Natl. Acad. Sci.* **95**, 15166–15168 (1998).
21. TF Varley, M Pope, null, null, O Sporns, Partial entropy decomposition reveals higher-order information structures in human brain activity. *Proc. Natl. Acad. Sci.* **120**, e2300888120 (2023).
22. T Berger, WB Levy, A mathematical theory of energy efficient neural computation and communication. *IEEE Transactions on Inf. Theory* **56**, 852–874 (2010).
23. ST Piantadosi, H Tily, E Gibson, Word lengths are optimized for efficient communication. *Proc. Natl. Acad. Sci.* **108**, 3526–3529 (2011).
24. B Lee, et al., Dissecting landscape art history with information theory. *Proc. Natl. Acad. Sci.* **117**, 26580–26590 (2020).
25. AJ Majda, B Gershgorin, Quantifying uncertainty in climate change science through empirical information theory. *Proc. Natl. Acad. Sci.* **107**, 14958–14963 (2010).
26. P Nurse, Life, logic, and information. *Nature* **454**, 424–426 (2008).
27. W Szpankowski, A Grama, Frontiers of science information: Shannon meets turing. *IEEE Comput.* pp. 32–42 (2018).
28. Y Bar-Hillel, R Carnap, Semantic information. *The Br. J. Philos. Sci.* **4**, 147–157 (1953).
29. C Shannon, The lattice theory of information. *Transactions IRE Prof. Group on Inf. Theory* **1**, 105–107 (1953).
30. L Floridi, Outline of a theory of strongly semantic information. *Minds Mach.* **14** (2004).
31. K Devlin, *Logic and Information.* (Cambridge University Press), (1991).
32. J Bao, et al., Towards a theory of semantic communication in *IEEE Network Science Workshop.* (IEEE Computer Society, Los Alamitos, CA, USA), pp. 110–117 (2011).
33. J Bao, et al., Towards a theory of semantic communication in *Extended Technical Report.* (2011).
34. J Liu, W Zhang, HV Poor, A rate-distortion framework for characterizing semantic information in *IEEE International Symposium on Information Theory.* (2021).
35. J Liu, S Shao, W Zhang, HV Poor, An indirect rate-distortion characterization for semantic sources: General model and the case of gaussian observation. *IEEE Transactions on Commun.* **70**, 5946–5959 (2022).
36. Y Shao, Q Cao, D Gündüz, A theory of semantic communication. *IEEE Transactions on Mob. Comput.* **23**, 12211–12228 (2022).
37. T Guo, Z Song, H Wu, Y Li, Rate–distortion limits for task-oriented compression with side information. *Entropy* **28**, 593 (2026).
38. PA Stavrou, M Kountouris, The role of fidelity in goal-oriented semantic communication: A rate distortion approach. *IEEE Transactions on Commun.* **71**, 3918–3931 (2023).
39. D Gündüz, et al., Beyond transmitting bits: Context, semantics, and task-oriented communications. *IEEE J. on Sel. Areas Commun.* **41**, 5–41 (2023).
40. K Niu, P Zhang, *A Mathematical Theory of Semantic Communication.* (Springer and Posts & Telecom Press), (2025).
41. B Juba, M Sudan, Universal semantic communication I in *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08. (Association for Computing Machinery, New York, NY, USA), p. 123–132 (2008).
42. O Goldreich, B Juba, M Sudan, A theory of goal-oriented communication. *J. ACM* **59**, article 8 (2012).
43. H Yu, JA Evans, LR Varshney, Information lattice learning. *J. Artif. Intell. Res.* **77**, 971–1019 (2023).
44. H Yu, LR Varshney, Semantic compression with information lattice learning. *2024 IEEE Int. Symp. on Inf. Theory Work. (ISIT-W)* pp. 1–6 (2024).
45. CE Shannon, Coding theorems for a discrete source with a fidelity criterion in *Claude E. Shannon: Collected Papers.* (Wiley-IEEE Press), pp. 325–350 (1993).
46. CE Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423, 623–656 (1948).
47. T Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall electrical engineering series. (Prentice-Hall), (1971).
48. SJ Russell, P Norvig, *Artificial Intelligence: A Modern Approach (4th Edition).* (Pearson), (2020).
49. Meta Platforms Inc., I-JEPA: The first AI model based on Yann LeCun's vision for more human-like AI (https://ai.meta.com/blog/yann-lecun-ai-model-i-jepa) (2023).
50. Centaur AI Institute, 4th Neuro-Symbolic Summer School (https://learning.centaurinstitute.org) (2025).
51. D Calanzone, S Teso, A Vergari, Logically consistent language models via neuro-symbolic integration in *Proceedings of the Thirteenth International Conference on Learning Representations.* (2025).
52. S Carrow, et al., Neural reasoning networks: Efficient interpretable neural networks with automatic textual explanations in *Proceedings of the AAAI Conference on Artificial Intelligence.* (2025).
53. LA Lastras, et al., Towards a unification of logic and information theory (2024) https://arxiv.org/abs/2301.10414.
54. ACC Yao, Some complexity questions related to distributive computing(preliminary report) in *Proceedings of the Eleventh annual ACM Symposium on Theory of Computing*, STOC '79. (Association for Computing Machinery, New York, NY, USA), p. 209–213 (1979).
55. A Orlitsky, J Roche, Coding for computing. *IEEE Transactions on Inf. Theory* **47**, 903–917 (2001).
56. R Bruno, U Vaccaro, NP-hardness and approximation algorithms for constrained entropy maximization. *IEEE Transactions on Inf. Theory* **72**, 832–843 (2026).
57. MR Ebrahimi, J Chen, A Khisti, Minimum entropy coupling with bottleneck in *Advances in Neural Information Processing Systems (NeurIPS 2024).* Vol. 37, pp. 59655–59688 (2024).
58. D Cox, J Little, D O'Shea, *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra.* (Springer), (2015).
59. Y Breitbart, H Hunt, D Rosenkrantz, On the size of binary decision diagrams representing boolean functions. *Theor. Comput. Sci.* **145**, 45–69 (1995).
60. D Mehta, V Raghavan, Decision tree approximations of boolean functions. *Theor. Comput. Sci.* **270**, 609–623 (2002).
61. R O'Donnell, M Saks, O Schramm, R Servedio, Every decision tree has an influential variable in *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05).* pp. 31–39 (2005).
62. D Slepian, J Wolf, Noiseless coding of correlated information sources. *IEEE Transactions on Inf. Theory* **19**, 471–480 (1973).
63. A Wyner, J Ziv, The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Inf. Theory* **22**, 1–10 (1976).
64. R Zamir, The rate loss in the wyner-ziv problem. *IEEE Transactions on Inf. Theory* **42**, 2073–2084 (1996).
65. P Elias, Universal codeword sets and representations of the integers. *IEEE Transactions on Inf. Theory* **21**, 194–203 (1975).
66. D Sivan, M Tsodyks, Information rate of meaningful communication. *Proc. Natl. Acad. Sci.* **122**, e2502353122 (2025).
67. T Cover, Enumerative source encoding. *IEEE Transactions on Inf. Theory* **19**, 73–77 (1973).
68. A Orlitsky, Interactive communication: balanced distributions, correlated files, and average-case complexity in *Proceedings 32nd Annual Symposium of Foundations of Computer Science.* pp. 228–238 (1991).
69. I Kontoyiannis, Pointwise redundancy in lossy data compression and universal lossy data compression. *IEEE Transactions on Inf. Theory* **46**, 136–152 (2000).
70. DJ MacKay, *Information Theory, Inference, and Learning Algorithms.* (Cambridge University Press), (2003).
71. N Immerman, *Descriptive Complexity.* (Springer, New York USA), (1999).
72. R Fagin, JY Hailpern, Y Moses, M Vardi, *Reasoning about Knowledge.* (MIT Press), (2004).
73. J Väänänen, Second-order and Higher-order Logic in *The Stanford Encyclopedia of Philosophy*, ed. EN Zalta. (Metaphysics Research Lab, Stanford University), Fall 2021 edition, (2021) https://plato.stanford.edu/archives/fall2021/entries/logic-higher-order/.
74. NJ Nilsson, Probabilistic logic. *Artif. Intell.* **28**, 71–87 (1986).
75. R Marinescu, et al., Logical credal networks in *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22. (2022).
76. R Fagin, R Riegel, A Gray, Foundations of reasoning with uncertainty via real-valued logics. *Proc. Natl. Acad. Sci.* **121** (2024).
77. P Grunwald, P Vitanyi, Shannon information and kolmogorov complexity (2004).