# The Rate-Distortion-Perception Trade-Off with Algorithmic Realism

**Yassine Hamdi**
Department of Electrical
and Electronic Engineering
Imperial College London, UK
`y.hamdi@imperial.ac.uk`

**Aaron B. Wagner**
School of Electrical
and Computer Engineering
Cornell University, USA
`wagner@cornell.edu`

**Deniz Gündüz**
Department of Electrical
and Electronic Engineering
Imperial College London, UK
`d.gunduz@imperial.ac.uk`

## Abstract

Realism constraints (or constraints on perceptual quality) have received considerable recent attention within the context of lossy compression, particularly of images. Theoretical studies of lossy compression indicate that high-rate common randomness between the compressor and the decompressor is a valuable resource for achieving realism. On the other hand, the utility of significant amounts of common randomness at test time has not been noted in practice. We offer an explanation for this discrepancy by considering a realism constraint that requires satisfying a universal critic that inspects realizations of individual compressed images, or batches thereof. We characterize the optimal rate-distortion-perception trade-off under such a realism constraint, and show that it is asymptotically achievable without any common randomness, unless the batch size is impractically large.

## 1 Introduction

Realism, or perceptual quality, of reconstructed signals is a long-standing open challenge in lossy compression, particularly for image/video compression [8, 25]. It has received renewed interest in the recent years due to the remarkable progress in image generation models and neural compression techniques. The idea is that reconstructed images should be indistinguishable to humans from naturally occurring ones in addition to having a high pixel-level fidelity to the original source. This ensures that reconstructed images are free of obvious artifacts such as blocking, blurriness, etc.

The idea that the output of the decoder should resemble the source in a statistical sense is not new. Advanced Audio Coding (AAC), for instance, includes a provision to add high-frequency noise to the output so that its power spectrum resembles that of the source [19]. But the idea has received renewed attention with the emergence of adversarial loss functions in learned compression [18, 23, 2, 3]. In practice, this has proven to be a powerful method for ensuring that reconstructed images have high perceptual quality [2, 15, 11, 13]. Adversarial loss functions can in many cases be viewed as variational forms of statistical divergences. Thus one can think of constraining the distribution of reconstructions to be close to that of the source according to some divergence, in addition to requiring that each reconstructed image be close to its respective source according to conventional notions of distortion.

Rate-distortion theory characterizes the optimal trade-off between rate and distortion in lossy compression [16, 19]. The fundamental object in the theory is the *rate-distortion function*, for a given source distribution $p_X$ :

$$\Delta \in [0, \infty) \mapsto R^{(0)}(\Delta) := \min_{p_{Y|X} \text{ s.t. } \mathbb{E}_p[d(X,Y)] \leq \Delta} I_p(X; Y), \tag{1}$$

where $p_{X,Y}$ is defined as $p_X \cdot p_{Y|X}$. This function has been shown to describe the optimal trade-off between rate and distortion under a variety of assumptions. Blau and Michaeli [3] postulated an

augmented form that includes a *distribution matching* constraint, which they call the *rate-distortion-perception* (RDP) function

$$(\Delta, \lambda) \in [0, \infty)^2 \mapsto R^{(1)}(\Delta, \lambda) := \min_{p_{Y|X} \text{ s.t. } \mathcal{D}(p_X, p_Y) \leq \lambda, \mathbb{E}_p[d(X,Y)] \leq \Delta} I_p(X; Y), \qquad (2)$$

where $\mathcal{D}$ can be any divergence between distributions. This function has likewise been shown to describe the optimal trade-off between rate, distortion, and realism under a variety of assumptions [22, 5]. Curiously, however, these results show that substantial amounts of high-quality common randomness are needed to meet the $R^{(1)}(\cdot, 0)$ bound [17, 24, 5] — see also [26]. The exception is the case in which the realism constraint is imposed in a very weak form, namely that the histograms of the source and reconstruction images should be close on a per-realization basis [5]. Note that common pseudorandomness, say generated from a shared seed, does not qualify as common randomness for the purposes of the above results.

On the other hand, the theoretical prediction that lossy compression schemes would benefit from substantial amounts of high-quality common randomness between the encoder and decoder has not been observed in practice. To the best of our knowledge, there exist compression schemes [1, 11, 12, 9, 15, 27], considered as state-of-the-art, that do not involve any common randomness. While it is possible that future designs will find common randomness to be a valuable resource, it seems more likely that the discrepancy between the theoretical prediction and practical experience lies with a flaw with the theoretical models.

Consider a communication system for which a strong realism constraint is imposed: the distribution of the reconstructions must be close to the distribution of natural images, say, in Wasserstein or total variation distance (TVD). If the source distribution is continuous, then the code cannot be deterministic, for otherwise the reconstruction distribution would be supported on a countable set (corresponding to the set of received bit strings). Thus some amount of randomization is required to meet the constraint. The decoder can randomize its output in a way that "spreads" the point masses out to form a continuous distribution, but adding independent noise at the decoder inevitably degrades the distortion. Common randomness is useful because it allows the discrete reconstruction points to be dispersed to form a continuous distribution without less overall distortion. This is the basis for the finding that common randomness is a useful resource for compression under realism constraints [21].[1]

The above reasoning is evidently sensitive to the nature of the realism constraint. If we simply require that each reconstructed image appear realistic in its own right, without reference to the reconstruction ensemble, then the spreading process mentioned above is unnecessary. It follows that there would be no need for randomization. This is relevant because human observers, who are the ultimate arbiters of realism in practice, are adept at identifying unrealistic features of individual images. Yet it is difficult for human observers to distinguish between a continuous ensemble and one that is discrete with a very large support set, since doing so would require viewing (and remembering) many images. In short, human critics are very good at spotting unrealistic aspects of individual images but are expected to be poor at detecting subtle ensemble-level differences.

This suggests posing the realism constraint in a way that better captures the relative strengths and weaknesses of human critics. The aforementioned strong realism constraint has also been challenged in the context of other problems, such as generative modeling [20]. We consider a novel formulation of the lossy compression problem in which the goal is to satisfy a critic that is incredibly discriminating when viewing individual images. In fact, a reconstructed image is declared unrealistic if there exists some computable test, no matter how complex, that can distinguish it from the set of typical source images (see Definition 3.5 to follow). At the same time, we assume that the critic can glean information about the ensemble only by inspecting batches of individual samples. Under this formulation, we show that the rate-distortion-perception function $R^{(1)}(\cdot, 0)$ in (2) is achievable without common randomness unless the batch size is unreasonably high—on par with the number of possible outputs of the decoder (Theorems 4.1 and C.1). If common randomness is not needed to fool this critic, it should not be needed to fool any weaker (and more practical) critic, since the stronger critic subsumes the weaker one. This is akin to how in cryptography one might prove security guarantees assuming a very strong adversary, stronger than can be implemented in practice. The fact that the adversary cannot be practically implemented is a strength of our approach. It is notable that

---

[1]It is now apparent why sharing a pseudorandom seed is insufficient, as this would expand the number of distinct reproductions by a multiplicative factor equal to the number of possible values of the seed, which is relatively small if the seed is short.

there exist compressors that can satisfy such discriminating critics at all. It is all the more notable that such critics can be satisfied while achieving the rate-distortion-perception function $R^{(1)}(\cdot, 0)$ in (2), which is the most optimistic rate-distortion trade-off possible under the circumstances. Conversely, we show that common randomness is indeed beneficial if the batch size is extremely large, larger than would ever occur in practice (Theorem 4.3). In this regime, our realism measure reduces to a divergence and common randomness is again useful. These two results clarify that common randomness is indeed useful, consistent with theoretical predictions, but only in regimes that do not occur in practice, consistent with the current state of the experimental literature. Our results show the existence of optimal schemes which do not involve any common randomness at test time, but there may exist other optimal schemes, which rely on common randomness at test time, as well as learned schemes relying on common randomness at training time.

In Section 2, we provide some background on the formalism for critics in algorithmic information theory. In Section 3, we introduce our new formalism for the RDP trade-off, including in asymptotic regimes where the source is of the form $p^{\otimes n}$, where $n \to \infty$. In Section 4, we present our main asymptotic results, which pertain to low batch size and large batch size regimes. Due to space limitations, achievable points in the one-shot setting are provided in Appendix C.

## 2 Background

The cardinality of a finite set $\mathcal{X}$ is denoted $|\mathcal{X}|$. We denote by $[a]$ the set $[1, a] \cap \mathbb{N}$. We use $x_{1:n}$ to denote a finite sequence $(x_1, ..., x_n)$, and $\mathbf{x}^{(n,b)}$ to denote a batch $\{x_{1:n}^{(k)}\}_{k \in [b]}$ of $b$ strings, each being of length $n$. We abbreviate $\mathbf{x}^{(1,b)}$ with $\mathbf{x}^{(b)}$. We use $\equiv$ to denote equality of distributions, and $I_p(X; Y)$ to denote the mutual information between $X$ and $Y$ with respect to joint distribution $p_{X,Y}$. We denote by $\{0, 1\}^*$ the set of non-empty finite strings of 0's and 1's. For any nonempty finite set $\mathcal{X}$, and any distribution $p$ on $\mathcal{X}$, we denote by $p^{\otimes *}$ the function defined on $\{0, 1\}^*$, which is null outside of $\cup_{n \in \mathbb{N}} \mathcal{X}^n$, and such that for every $n \in \mathbb{N}$, the restriction of $p^{\otimes *}$ on $\mathcal{X}^n$ is $p^{\otimes n}$. The total variation distance between distributions $p$ and $q$ is denoted $\|p - q\|_{TV}$.

### 2.1 Background on algorithmic information theory

The theory of $p$-critics and universal critics has recently been brought to the attention of the machine vision community via [20]. We refer to it for readers interested in a high-level and insightful presentation of the topic and its usefulness in diverse machine learning tasks (generative modeling, outlier detection). Relevant background on computability theory is provided in Appendix B. Throughout the paper, we assume that the source $X$ follows a distribution $p_X$ on a finite set $\mathcal{X}$, and that $p_X$ is a computable function from $\mathcal{X}$ to $(0, 1)$. We identify every element of $\mathcal{X}$ with a string of 0's and 1's, via an injection from $\mathcal{X}$ to $\{0, 1\}^s$, for some $s \in \mathbb{N}$. For example, if $\mathcal{X}$ is a set of images of a given resolution, then one can identify each image with the corresponding output from a fixed-length lossless compressor. The following definition is substantially close to Definition 4.3.8 in [14]. See also Lemma 4.3.5 in [14].

**Definition 2.1.** *Consider a finite set $\mathcal{X}$, identified with a subset of $\{0, 1\}^s$. Let $p$ be a distribution on $\mathcal{X}$ such that $\forall x \in \mathcal{X}, p(x) > 0$. A p-critic is a function $\delta : \mathcal{X} \to \mathbb{R}$, such that*

$$\sum_{x \in \mathcal{X}} p(x) 2^{\delta(x)} \leq 1. \tag{3}$$

*A $p^{\otimes *}$ -critic is a function $\delta : \cup_{n \in \mathbb{N}} \mathcal{X}^n \to \mathbb{R}$, such that for every input dimension $n \in \mathbb{N}$, we have*

$$\sum_{x \in \mathcal{X}^n} p^{\otimes n}(x) 2^{\delta(x)} \leq 1. \tag{4}$$

The notion of $p^{\otimes *}$-critic in Definition 2.1 is used to study an asymptotic regime in Section 3.2. Note that for any probability distribution $\pi$ on $\mathbb{N}$, the mixture $\tilde{p} := \sum_{n \in \mathbb{N}} \pi(n) p^{\otimes n}$ is a probability measure. By multiplying (4) by $\pi_n$, and summing over $n$, we obtain

$$\sum_{x \in \cup_{n \in \mathbb{N}} \mathcal{X}^n} \tilde{p}(x) 2^{\delta(x)} \leq 1. \tag{5}$$

3
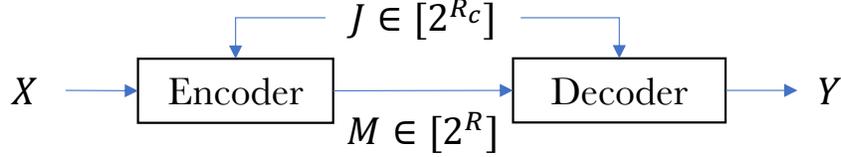
Figure 1: The system model for the one-shot setting.

Hence, a $p$-critic (resp. $p^{\otimes*}$-critic) is akin to a log-likelihood ratio: given a $p$-critic (resp. $p^{\otimes*}$-critic) $\delta$, setting $q : x \mapsto p(x)2^{\delta(x)}$ (resp. $q : x \mapsto \tilde{p}(x)2^{\delta(x)}$) gives

$$\forall x \in \mathcal{X} \text{ s.t. } p(x) > 0, \ \delta(x) = \log\left(\frac{q(x)}{p(x)}\right) \text{ (resp. } \log\left(\frac{q(x)}{\tilde{p}(x)}\right)), \quad \text{and} \quad \sum_{x \in \mathcal{X}} q(x) \leq 1. \quad (6)$$

Links to hypothesis testing are discussed in [20], where a sample $x$ is deemed unrealistic if the likelihood ratio is large enough. Hence, intuitively, $\delta(x)$ can be considered as a measure of *realism deficiency* of $x$. The strength of this theory lies in the existence of objects (critics, measures) having a so-called *universality property*. For the purpose of clarity, we defer such results to Appendix B, as they are only used in our proofs.

## 3 New model for the rate-distortion-perception trade-off

### 3.1 The one-shot setting

We consider a function $d : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ called the distortion measure. A compression scheme can be randomized, and potentially leverage available common randomness $J$ between the encoder and the decoder, as depicted in Figure 1 and formalized in the following definition.

**Definition 3.1.** *Given non-negative reals $R$ and $R_c$, an $(R, R_c)$ code is a privately randomized encoder and decoder couple $(F, G)$ consisting of a conditional distribution $F_{M|X,J}$ from $\mathcal{X} \times [2^{R_c}]$ to $[2^R]$, and a conditional distribution $G_{Y|M,J}$ from $[2^R] \times [2^{R_c}]$ to $\mathcal{X}$. Variables $M$ and $Y$ are called the message and reconstruction, respectively, and distribution*

$$P := p_X \cdot p^{\mathcal{U}}_{[2^{R_c}]} \cdot F_{M|X,J} \cdot G_{Y|M,J}$$

*is called the distribution induced by the code. Moreover, such a code is said to be deterministic if $R_c = 0$ and mappings $F, G$ are deterministic.*

We propose a new RDP trade-off, as follows.

**Definition 3.2.** *We extend $d$ into an additive distortion measure on batches of elements of $\mathcal{X}$: for all $B \in \mathbb{N}$,*

$$\forall (\mathbf{x}^{(B)}, \mathbf{y}^{(B)}) \in \mathcal{X}^B \times \mathcal{X}^B, \quad d(\mathbf{x}^{(B)}, \mathbf{y}^{(B)}) := \tfrac{1}{B}\sum_{k=1}^{B} d(x^{(k)}, y^{(k)}).$$

**Definition 3.3.** *Consider a positive integer $B$, and a $p_X^{\otimes B}$-critic $\delta$. A tuple $(R, \Delta, C)$ is said to be $\delta$-achievable with algorithmic realism if there exists some $R_c \in \mathbb{R}_{\geq 0}$ and an $(R, R_c)$ code such that the distribution $P$ induced by the code satisfies*

$$\mathbb{E}_{P^{\otimes B}}\left[d(\mathbf{X}^{(B)}, \mathbf{Y}^{(B)})\right] \leq \Delta \ and \tag{7}$$

$$\mathbb{E}_{P^{\otimes B}}[\delta(\mathbf{Y}^{(B)})] \leq C, \tag{8}$$

*where $\mathbf{X}^{(B)}$ denotes a batch of $B$ i.i.d. source samples, and $\mathbf{Y}^{(B)}$ the batch of corresponding reconstructions produced by the code (with each source sample being compressed separately). If the code is deterministic, then we say that $(R, \Delta, C)$ is $\delta$-achievable with a deterministic code.*

Due to space limitations, we provide achievable points in Appendix C. In the next section, we define an asymptotic notion of achievability.

4
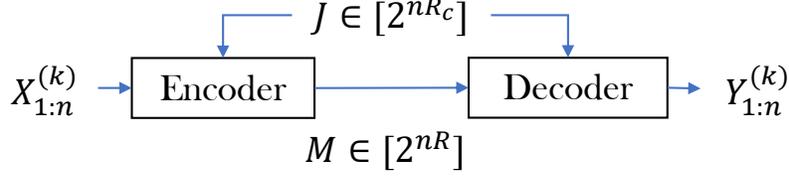
Figure 2: The system model for the asymptotic setting. Index $k$ ranges from 1 to the batch size. The same encoder-decoder pair is used to process each source sample in the batch.

## 3.2 Asymptotic setting

In order to derive insight into the corresponding RDP trade-off, we study a special case, which is typical in the information theory literature. We consider the compression of a source distributed according to $p_X^{\otimes n}$, with $n$ a large integer. More precisely, we characterize the RDP trade-off in asymptotic settings where both $n$ and the batch size go to infinity. The extension of $d$ into an *additive distortion measure* on finite sequences, and batches of finite sequences, follows from Definition 3.2. The setup is depicted in Figure 2. Given a coding scheme, each item in a batch of source samples is compressed separately, and realism is measured based on the resulting batch of reconstructions. This is formalized in the definition below.

**Definition 3.4.** *Given $R, R_c \geq 0$, and $n \in \mathbb{N}$, a $(n, R, R_c)$ code is a privately randomized encoder and decoder couple $(F^{(n)}, G^{(n)})$ consisting of a mapping $F^{(n)}_{M|X_{1:n}, J}$ from $\mathcal{X}^n \times [2^{nR_c}]$ to $[2^{nR}]$ and a mapping $G^{(n)}_{Y_{1:n}|M, J}$ from $[2^{nR_c}] \times [2^{nR}]$ to $\mathcal{X}^n$. Moreover, such a code is said to be fully deterministic if $R_c = 0$ and both $F^{(n)}$ and $G^{(n)}$ are deterministic. The distribution induced by the code is*

$$P^{(n)} := p_X^{\otimes n} \cdot p_{[2^{nR_c}]}^{\mathcal{U}} \cdot F^{(n)}_{M|X_{1:n}, J} \cdot G^{(n)}_{Y_{1:n}|M, J},$$

*and variable $Y_{1:n}$ is called the reconstruction.*

We define asymptotic achievability as follows. See Appendix B for background on computability.

**Definition 3.5.**
*A quadruplet $(R, R_c, \{B_n\}_{n \geq 1}, \Delta)$ is said to be asymptotically achievable with algorithmic realism if for any $\varepsilon > 0$, there exists a sequence of codes $\{(F^{(n)}, G^{(n)})\}_n$, the $n$-th being $(n, R + \varepsilon, R_c)$, such that the sequence $\{P^{(n)}\}_n$ of distributions induced by the codes satisfies*

$$\limsup_{n \to \infty} \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \left[ d(\mathbf{X}^{(n, B_n)}, \mathbf{Y}^{(n, B_n)}) \right] \leq \Delta + \varepsilon, \tag{9}$$

*and for any lower semi-computable $p_X^{\otimes *}$-critic $\delta$,*

$$\sup_{n \in \mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \left[ \delta(\mathbf{Y}^{(n, B_n)}) \right] < \infty. \tag{10}$$

*We say that $(R, \{B_n\}_{n \geq 1}, \Delta)$ is achievable with a fully deterministic scheme if for each $n$, the code $(F^{(n)}, G^{(n)})$ is fully deterministic.*

Constraint (10) is very stringent: a single compression scheme is to satisfy a performance guarantee for every lower semi-computable $p_X^{\otimes *}$-critic (i.e. every relevant one). The motivation for the specific form of (10) is firstly from the algorithmic information theory literature: it is stated in [14] (p.140) that a sample from a large set, identified to a long string of 0's and 1's of some length $k$, is realistic if its realism deficiency is small compared to $k$. The constraint in (10) is at least as stringent, since in our asymptotic setting, each $x_{1:n} \in \mathcal{X}^n$ is identified with a string of length linear in $n$, while we require the realism deficiency to be bounded. Moreover, consider the following simple example. Assume $\mathcal{X} = \{0, 1\}$, and $p_X$ is a Bernoulli distribution $B(\rho)$. Consider the 0-1 distortion (also called Hamming distortion), and some distortion level $\Delta < \min(\rho, 1 - \rho)$. Then, for large enough $n$, the classical rate-distortion optimal code appearing in the information theory literature produces reconstructions having a frequency of 1's of roughly $(\rho - \Delta)/(1 - 2\Delta)$ ([6], Sections 10.3.1 and 10.5), i.e. different from $\rho$ (if $\rho \neq 1/2$ and $\Delta > 0$). Then, for the $p_X^{\otimes *}$-critic appearing in Appendix I (Claim I.1), which involves the frequency of occurrence of a pattern, the expected score diverges as $n$ goes to

5

infinity. Hence, constraint (10) is not satisfied by such a code, optimized only for rate and distortion, but not for realism. This concludes the definitions for our setup. In the next sections, we present our results in asymptotic settings. See Appendix C for achievable points in the one-shot setting.

# 4 Results

## 4.1 Low batch size regime

The following theorem states that $R^{(1)}(\cdot, 0)$, defined in (2), which naturaly arises in the distribution matching formalism, also characterizes the optimal trade-off in our asymptotic setting, when the batch size is not impractically large.

**Theorem 4.1.** *Consider a computable sequence $\{B_n\}_{n\geq 1}$ of positive integers such that $\log(B_n)/n \to 0$. For any $\Delta \in \mathbb{R}_+$, let $R(\Delta)$ be the infimum of rates $R$ such that there exists $R_c \in \mathbb{R}_{\geq 0}$ such that $(R, R_c, \{B_n\}_{n\geq 1}, \Delta)$ is asymptotically achievable with algorithmic realism. Moreover, for any $\Delta \in \mathbb{R}_+$, let $R_*(\Delta)$ be the infimum of rates $R$ such that $(R, \{B_n\}_{n\geq 1}, \Delta)$ is asymptotically achievable with algorithmic realism with fully deterministic codes. Then, we have*

$$\forall \Delta \in \mathbb{R}_+ \text{ s.t. } R^{(1)}(\Delta, 0) < H_p(X), \text{ we have } R(\Delta) = R_*(\Delta) = R^{(1)}(\Delta, 0). \qquad (11)$$

The proof is provided in Appendices E and F. The strength of this result lies in how stringent constraint (10) is: a single compression scheme satisfies a performance guarantee for every relevant $p_X^{\otimes *}$-critic, and deterministic schemes are sufficient. The proof leverages the existence of a *universal* $p_X^{\otimes *}$-critic $\delta_0$ (see Appendix B.2). Indeed, it is sufficient to construct a scheme which achieves (10) only for such a $\delta_0$, which is more sensitive than all relevant $p_X^{\otimes *}$-critics. It is a very strong critic, stronger than can be implemented in practice, which is another strength of Theorem 4.1.

## 4.2 Generalizing the distribution matching formalism

Under the distribution matching formalism for the RDP trade-off, the natural asymptotic notion of achievability is as follows.

**Definition 4.2.** *[17, 3]*
*A quadruplet $(R, R_c, \{B_n\}_{n\geq 1}, \Delta)$ is said to be asymptotically achievable with near-perfect realism if for any $\varepsilon > 0$, there exists a sequence of codes $\{(F^{(n)}, G^{(n)})\}_n$, the $n$-th being $(n, R + \varepsilon, R_c)$, such that the sequence $\{P^{(n)}\}_n$ of distributions induced by the codes satisfies*

$$\limsup_{n\to\infty} \mathbb{E}_{P^{(n)}}\left[d(X_{1:n}, Y_{1:n})\right] \leq \Delta + \varepsilon \quad and \quad \|P_{Y_{1:n}}^{(n)} - p_X^{\otimes n}\|_{TV} \xrightarrow[n\to\infty]{} 0. \qquad (12)$$

The total variation distance (TVD) in (12) is directly related to the performance of the optimal hypothesis tester between the reconstruction distribution $P_{Y_{1:n}}^{(n)}$, and the source distribution $p_X^{\otimes n}$ [3]. Replacing the second property in (12) with

$$\exists N \in \mathbb{N}, \forall n \geq N, \ P_{Y_{1:n}}^{(n)} \equiv p_X^{\otimes n} \qquad (13)$$

gives the notion of asymptotic *achievability with perfect realism*. It was shown that these two notions are equivalent for finite-valued sources [17], as well as for continuous sources under mild assumptions [17, 24]. As formalized in the theorem below, in a certain large batch size regime, asymptotic achievability with algorithmic realism is equivalent to asymptotic achievability with near-perfect realism. The proof is provided in Appendix G.

**Theorem 4.3.** *Consider a computable increasing sequence $\{B_n\}_{n\geq 1}$ in $\mathbb{N}$ such that $B_n/|\mathcal{X}|^n \to \infty$. Then, for any $R_c \in \mathbb{R}_{\geq 0}$, and any $(R, \Delta) \in (\mathbb{R}_+)^2$, tuple $(R, R_c, \{B_n\}_{n\geq 1}, \Delta)$ is asymptotically achievable with algorithmic realism if and only if $(R, R_c, \Delta)$ is asymptotically achievable with near-perfect realism, if and only if $(R, R_c, \Delta)$ is asymptotically achievable with perfect realism.*

Hence, Theorem 4.3, similarly to the finding in [20], shows that for large batch size, our formalism is equivalent to the distribution matching formalism. Hence, the former is a generalization of the latter. Moreover, Theorem 4.3 implies that common randomness is useful when the size of the batch inspected by the critic is extremely large. As discussed in Section 4.1, Theorem 4.1 states that common randomness does not improve the trade-off under our formalism, in all regimes where the batch size is not impractically large with respect to the dimension $n$ of the source. Thus, Theorems 4.1 and 4.3 show that, in order to understand the role of randomization in lossy compression with realism constraints, the focus should be shifted to the size of the batch inspected by the critic.

## Acknowledgments and Disclosure of Funding

## References

[1] E. Agustsson, D. Minnen, G. Toderici, and F. Mentzer. Multi-Realism Image Compression with a Conditional Generator. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[2] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. Van Gool. Generative Adversarial Networks for Extreme Learned Image Compression. In *IEEE/CVF International Conference on Computer Vision*, 2019.

[3] Y. Blau and T. Michaeli. Rethinking Lossy Compression: The Rate-Distortion-Perception Tradeoff. In *36th International Conference on Machine Learning*, 2019.

[4] C. L. Canonne. A short note on learning discrete distributions, 2020. arxiv:2002.11457.

[5] J. Chen, L. Yu, J. Wang, W. Shi, Y. Ge, and W. Tong. On the Rate-Distortion-Perception Function. *IEEE Journal on Selected Areas in Information Theory*, 3(4), 2022.

[6] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley-Interscience. Wiley, 2006.

[7] P. Cuff. Distributed Channel Synthesis. *IEEE Transactions on Information Theory*, 59(11), 2013.

[8] M. P. Eckert and A. P. Bradley. Perceptual quality metrics applied to still image compression. *Signal Processing*, 70(3):177–200, 1998.

[9] N. F. Ghouse, J. Petersen, A. Wiggers, T. Xu, and G. Sautière. A Residual Diffusion Model for High Perceptual Quality Codec Augmentation, 2023. arxiv:2301.05489.

[10] Y. Hamdi, A. B. Wagner, and D. Gündüz. The Rate-Distortion-Perception Trade-off: the Role of Private Randomness. In *IEEE International Symposium on Information Theory (ISIT)*, 2024.

[11] D. He, Z. Yang, H. Yu, T. Xu, J. Luo, Y. Chen, C. Gao, X. Shi, H. Qin, and Y. Wang. PO-ELIC: Perception-Oriented Efficient Learned Image Coding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2022.

[12] E. Hoogeboom, E. Agustsson, F. Mentzer, L. Versari, G. Toderici, and L. Theis. High-fidelity image compression with score-based generative models, 2023. arXiv:2305.18231.

[13] S. Iwai, T. Miyazaki, and S. Omachi. Controlling rate, distortion, and realism: Towards a single comprehensive neural image compression model. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024.

[14] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Texts in Computer Science. Springer International Publishing, Cham, 4th edition, 2019.

[15] F. Mentzer, G. D. Toderici, M. Tschannen, and E. Agustsson. High-Fidelity Generative Image Compression. *Advances in Neural Information Processing Systems*, 33, 2020.

[16] W. A. Pearlman and A. Said. *Digital Signal Compression: Principles and Practice*. Cambridge University Press, Cambridge (England), 2011.

[17] N. Saldi, T. Linder, and S. Yüksel. Output Constrained Lossy Source Coding With Limited Common Randomness. *IEEE Transactions on Information Theory*, 61(9), 2015.

[18] S. Santurkar, D. Budden, and N. Shavit. Generative Compression. In *Picture Coding Symposium*, 2018.

[19] K. Sayood. *Introduction to Data Compression*. Morgan Kaufmann, Waltham, MA (United States of America), 4th edition, 2012.

[20] L. Theis. Position: What makes an image realistic? In *Forty-first International Conference on Machine Learning*, 2024.

[21] L. Theis and E. Agustsson. On the advantages of stochastic encoders. In *Neural Compression: From Information Theory to Applications – workshop at the International Conference on Learning Representations*, 2021.

[22] L. Theis and A. B. Wagner. A coding theorem for the rate-distortion-perception function. In *Neural Compression: From Information Theory to Applications – workshop at the International Conference on Learning Representations 2021*.

[23] M. Tschannen, E. Agustsson, and M. Lucic. Deep Generative Models for Distribution-Preserving Lossy Compression. In *NeurIPS*, 2018.

[24] A. B. Wagner. The Rate-Distortion-Perception Tradeoff: The Role of Common Randomness, 2022. arXiv:2202.04147.

[25] H. R. Wu, W. Lin, and L. J. Karam. An overview of perceptual processing for digital pictures. In *IEEE International Conference on Multimedia and Expo Workshops*, 2012.

[26] T. Xu, Q. Zhang, Y. Li, D. He, Z. Wang, Y. Wang, H. Qin, Y. Wang, J. Liu, and Y. Zhang. Conditional perceptual quality preserving image compression, 2023. arXiv:2308.08154.

[27] R. Yang and S. Mandt. Lossy Image Compression with Conditional Diffusion Models. In *NeurIPS*, 2023.

# A Additional notation

Calligraphic letters such as $\mathcal{X}$ denote sets, except in $p_{\mathcal{J}}^{\mathcal{U}}$, which denotes the uniform distribution over set $\mathcal{J}$. Given a real number $\tau$, we denote by $\lfloor \tau \rfloor$ (resp. $\lceil \tau \rceil$) the largest (resp. smallest) integer less (resp. greater) than or equal to $\tau$. The length of a string $x$ is denoted by $l(x)$. We denote the set of (strictly) positive reals by $\mathbb{R}_+$, the set of (strictly) positive integers by $\mathbb{N}$, the set of rational numbers by $\mathbb{Q}$, and the Borel $\sigma$-algebra of $\mathbb{R}$ by $\mathcal{B}(\mathbb{R})$. The closure of a set $\mathcal{A}$ is denoted by $cl(\mathcal{A})$. Logarithms are in base 2. For a finite set $\mathcal{X}$, the empirical distribution of a sequence $x_{1:n} \in \mathcal{X}^n$ is denoted $\mathbb{P}_{\mathcal{X}}^{\mathrm{emp}}(x_{1:n})$. Given a distribution $P_{X_{1:n}}$ on $\mathcal{X}^n$, we denote by $\hat{P}_{\mathcal{X}}[X_{1:n}]$ the *average marginal distribution* of random string $X_{1:n}$, i.e., the distribution on $\mathcal{X}$ defined by:

$$\hat{P}_{\mathcal{X}}[X_{1:n}] := \tfrac{1}{n} \sum_{t=1}^{n} P_{X_t}.$$

# B Further background on algorithmic information theory

## B.1 Computability

This definition matches Definition 1.7.4 in [14], except for the definition of a computable real number, which we adapted from Exercise 1.7.22 of [14], and for the definition of a computable set, which matches that of [14], page 32.

**Definition B.1.** *Consider a subset $\mathcal{E}$ of $\mathbb{N}_{\geq 0}$. A map $f$ from $\mathcal{E}$ into $\mathbb{N}_{\geq 0}^3$ is said to be computable if it corresponds to a Turing machine (Section 1.7.1 in [14]). This notion extends to functions having as domain other common countable sets, such as $\mathbb{N}_{\geq 0}^k$ for $k \in \mathbb{N}$, and $\{0,1\}^*$, or any subset thereof, by identifying elements of these sets with non-negative integers via some reference bijections. Consider a computable map $f$ from a subset $\mathcal{E}$ of $\mathbb{N}_{\geq 0}$ into $\{0,1\} \times \mathbb{N}_{\geq 0} \times \mathbb{N}$. Then, composing $f$ with $(s, a, b) \mapsto (2s-1)a/b$ yields a map from $\bar{\mathcal{E}}$ to $\mathbb{Q}$, which is said to be a computable map from $\mathcal{E}$ to $\mathbb{Q}$. A map $f$ from a subset $\mathcal{E}$ of $\mathbb{N}_{\geq 0}$ into $\mathbb{R}$ is said to be lower semi-computable if there exists a computable function $\varphi$ from $\mathcal{E} \times \mathbb{N}$ into $\mathbb{Q}$, such that*

$$\forall x \in \mathcal{E}, \varphi(x, k) \underset{k \to \infty}{\to} f(x), \qquad \text{and} \qquad \forall x \in \mathcal{E}, \forall k \in \mathbb{N}, \ \varphi(x, k+1) \geq \varphi(x, k).$$

*Moreover, $f$ is said to be a computable map from $\mathcal{E}$ to $\mathbb{R}$ if both $f$ and $-f$ are lower semi-computable. A real number $\lambda$ is said to be computable if the constant function $f : \mathbb{N}_{\geq 0} \to \mathbb{R}$, $n \mapsto \lambda$ is a computable function from $\mathbb{N}_{\geq 0}$ to $\mathbb{R}$. A (possibly infinite) subset $\mathcal{X}$ of $\mathbb{N}_{\geq 0}$, is said to be computable if there exists a computable function $f$ from $\mathbb{N}_{\geq 0}$ to $\{0,1\}$, which returns 1 if its input is in $\mathcal{X}$, and 0 otherwise.*

The following lemma allows to construct (semi-)computable functions. Its proof is deferred to Appendix M.

**Lemma B.2.** *Let $\mathcal{E}$ denote a non-empty subset of $\mathbb{N}_{\geq 0}$, and let $f$ and $g$ denote functions from $\mathcal{E}$ to $\mathbb{R}$. (i) If $f$ and $g$ are both lower semi-computable, then functions $f + g$, $\lceil f \rceil$, and $2^f$ are lower semi-computable. If, in addition, $f$ and $g$ only take non-negative values, then $fg$ and $2^f/(3+f)^2$ are lower semi-computable. If, in addition, $f$ only takes positive values, then $\log(f)$ is lower semi-computable. (ii) If $f$ and $g$ are both computable, then functions $f + g$, $fg$, and $|f|$ are computable. If, in addition, $f$ only takes positive values, then functions $1/f$, and $f^{1/b}$ are computable, for any positive integer $b$. (iii) Let $\mathcal{X}$ be a computable finite subset of $\{0,1\}^*$. If $f$ is a lower semi-computable function from $\{0,1\}^*$ into $\mathbb{R}$, then the function $\tilde{f} : \{0,1\}^* \to \mathbb{R}$ which is null outside of $\cup_{n \in \mathbb{N}} \mathcal{X}^n$, and is defined by*

$$\forall x \in \cup_{n \in \mathbb{N}} \mathcal{X}^n, \ \tilde{f}(x) = \sum_{y \in \mathcal{X}^{l(x)}} f(y),$$

*is lower semi-computable. Moreover, if $p$ is a lower semi-computable probability measure on $\mathcal{X}$, then $p^{\otimes*}$ is lower semi-computable.*

## B.2 Universal critics and semi-measures

**Definition B.3.** *Given a finite set $\mathcal{W}$, a function $f : \mathcal{W} \to [0,1]$ is a semi-measure if*

$$\sum_{w \in \mathcal{W}} f(w) \leq 1.$$

*It is said to be a lower semi-computable semi-measure if $f$ is a semi-measure and $f$ is lower semi-computable.*

The following theorem, corresponds to Definition 4.3.2, Equation (4.2), and Theorems 4.3.1 and 4.3.3 in [14]. It introduces the notion of *universal $p^{\otimes*}$-critic*, used in [20]. The mixture $\mathbf{m}$ therein can be used as a prior distribution, which has been shown to be relevant in machine learning applications involving realism, such as outlier detection and generative modeling [20].

**Theorem B.4.** *Consider a finite set $\mathcal{X}$, each element of which is identified with a string in $\{0,1\}^s$, for some $s \in \mathbb{N}$. Let $p$ be a computable distribution on $\mathcal{X}$ such that $\forall x \in \mathcal{X}, p(x) > 0$. Then, there exists a $p^{\otimes*}$-critic $\delta_0$ (which is not necessarily lower semi-computable), such that for any lower semi-computable $p^{\otimes*}$-critic $\delta$, there exists a constant $c_\delta$ such that*

$$\forall x \in \bigcup_{n \in \mathbb{N}} \mathcal{X}^n, \quad \delta_0(x) \geq \delta(x) - c_\delta. \tag{14}$$

*Any $p^{\otimes*}$-critic satisfying (14) is called a universal $p^{\otimes*}$-critic.*

Since our definitions are slightly different from the classical ones, we provide a proof of Theorem B.4 in Appendix L.

Such a critic $\delta_0$ is one of the best measures of realism deficiency according to $p$, in the limit of arbitrarily long strings. If a critic $\delta$ identifies a certain amount of deficiency in a given string, then $\delta_0$ will identify at least as much deficiency, up to some additive constant. Intuitively, $\delta_0$ is sensitive to all properties of randomness according to $p$.

**Remark B.5.** *([14], Theorem 4.3.3) The universal semi-measure $\mathbf{m}$ can be chosen in such a way that*
$$\forall x \in \{0,1\}^*, \quad |-\log(\mathbf{m}(x)) - K(x)| \leq c, \tag{15}$$
*for some constant $c$, where $K$ is the Kolmogorov complexity (see, e.g., [14], Section 3.1). Property (15) constitutes a strong result, since the Kolmogorov complexity is only defined up to a constant -we omit the corresponding details, for the purpose of clarity. The map $x \mapsto \log(1/p(x)) - K(x)$ is sometimes considered to be an approximation of a universal $p^{\otimes*}$ critic, see, e.g., [20], and Appendix L.*

## C  One-shot achievable points

The quantities in this section should be understood as follows:

- $\mathcal{X}$ is a finite set of images, e.g. the set of all images of a given resolution, with a finite range for pixels (finite precision).
- $d$ is the mean squared error between pixel values.
- $B$ is the number of images inspected by the critic at a time.
- $R_1$ is the number of bits into which a given image is compressed.

No statistical assumption is made regarding the source distribution $p_X$ of images in $\mathcal{X}$. We have the following family of achievable points, in the sense of Definition 3.3.

**Theorem C.1.** *Consider a finite set $\mathcal{X}$ such that $|\mathcal{X}| \geq 2$, a computable distribution $p_X$ on $\mathcal{X}$ such that $\forall x \in \mathcal{X}, p_X(x) > 0$, a positive integer $B$, some $R > \log(B)/\log(\mathcal{X})$, some $\Delta \in \mathbb{R}_+$, and a $p_X^{\otimes B}$-critic $\delta$. Consider any conditional transition kernel $p_{Y|X}$ from $\mathcal{X}$ to $\mathcal{X}$ satisfying*

$$p_Y \equiv p_X, \quad \mathbb{E}_p[d(X,Y)] \leq \Delta. \tag{16}$$

*Then, for any $\varepsilon \in (0, \Delta/2)$, and any $\gamma > 0$, the triplet $(R_1, \Delta_1, C_1)$ is $\delta$-achievable, with a $(R_1, 0)$ code, where*

$$R_1 := R \log(|\mathcal{X}|) \tag{17}$$

$$\Delta_1 := \Delta + \varepsilon + \frac{6\Delta}{\varepsilon} \max(d) \cdot \eta_{R,\gamma} \tag{18}$$

$$C_1 := \frac{3\Delta}{\varepsilon} \Big[ \frac{B^2}{\lfloor 2^{R_1} \rfloor} + 2B\eta_{R,\gamma} \Big] \cdot \max_x B \log \frac{1}{p_X(x)} \tag{19}$$

$$\eta_{R,\gamma} := p(\mathcal{A}_{R,\gamma}) + 2^{-\gamma \log(|\mathcal{X}|)/2} \tag{20}$$

$$\mathcal{A}_{R,\gamma} := \Big\{ (x,y) \in \mathcal{X}^2 \mid \log \Big( \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \Big) - \log(\lfloor 2^{R_1} \rfloor) > -\gamma \log(|\mathcal{X}|) \Big\}, \tag{21}$$

*with the convention $0/0 := 1$.*

The proof is provided in Appendix D. This result provides insights on the asymptotic regime of Theorem 4.1. Consider the limit of large $|\mathcal{X}|$, with fixed $R, \Delta, \varepsilon, \gamma$, and with $\log(B) = o(\log|\mathcal{X}|)$. We know that

$$\mathbb{E}_p\Big[\log\Big(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\Big)\Big] = I_p(X;Y). \tag{22}$$

Hence, if this log-likelihood ratio concentrates well, and if $R_1 > I_p(X;Y)$, as in the definition of $R^{(1)}(\cdot, 0)$ in (2), then $p(\mathcal{A}_{R,\gamma})$ is small for small enough $\gamma$. In such an asymptotic regime, we obtain $\Delta_1 \approx \Delta$, and $C_1 = O(1)$. The assumption in Theorem 4.1, that the source is of the form $p_X^{\otimes n}$ for some large $n$, is only used to prove the concentration of this likelihood ratio. Hence, that result can be extended to a larger set of sources. The term $B^2/\lfloor 2^{R_1} \rfloor$ is an upper bound on the probability that two source samples in the batch are compressed into the same message. This is related to the so-called *birthday paradox* (see Appendix K). The term $\max_x B \log(1/p_X(x))$ is an upper bound on the output of $\delta$.

# D  Proof of Theorem C.1

## D.1  Outline

To show the achievability of a tuple $(R_1, \Delta_1, C_1)$, it is not necessary to construct an explicit compression scheme: it is sufficient to prove the abstract existence of such a scheme. To that end, we consider a set of random reconstructions, and study its realism properties in Section D.2. Then, we show the existence of a suitable choice of realizations of the latter reconstructions in Section D.3. In Section D.4, we prove Theorem C.1 by proposing a compression scheme achieving a close-to-uniform sampling from the set of reconstructions. For the remainder of Section D, we fix a finite set $\mathcal{X}$ such that $|\mathcal{X}| \geq 2$, a distribution $p_X$ on $\mathcal{X}$ such that $\forall x \in \mathcal{X}, p_X(x) > 0$, a positive integer $B$, and a $p_X^{\otimes B}$-critic $\delta$.

## D.2  Realism performance of a uniformly sampled batch of random reconstructions

### D.2.1  Random candidate reconstructions

Given a positive real $R_1$, let $\mathcal{C}$ be a family of $\lfloor 2^{R_1} \rfloor$ i.i.d. variables, each sampled from $p_X$. The $m$-th variable is denoted $y(\mathcal{C}, m)$. We denote their joint distribution by $Q_{\mathcal{C}}$. Given a realization $c$ of $\mathcal{C}$, we consider a batch $\mathbf{y}^{(B)}$ of $B$ elements of $c$, sampled uniformly with replacement. Then, we compute the batch's realism score $\delta(\mathbf{y}^{(B)})$. This is formalized in the following lemma, which gives an upper bound of the expected score with respect to $Q_{\mathcal{C}}$.

**Lemma D.1.** *Consider a positive real $R_1 \in (\log(B), \infty)$, and the following pmf.*

$$Q_{\mathcal{C},\mathbf{M}^{(B)},\mathbf{Y}^{(B)}}\big(\{y(m')\}_{m' \in \lfloor\!\lfloor 2^{R_1} \rfloor\!\rfloor}, \mathbf{m}^{(B)}, \mathbf{y}^{(B)}\big)$$

$$:= \Big(\prod_{m'=1}^{\lfloor 2^{R_1} \rfloor} p_X(y(m'))\Big) \cdot \frac{1}{\lfloor 2^{R_1} \rfloor^B} \cdot \prod_{k=1}^{B} \mathbf{1}_{y^{(k)}=y(m^{(k)})}. \tag{23}$$

*Then, we have*

$$\mathbb{E}_Q[\delta(\mathbf{Y}^{(B)})] \leq \frac{B^2}{\lfloor 2^{R_1} \rfloor} \max_x B \log \frac{1}{p_X(x)}. \tag{24}$$

The remainder of Section D.2 is dedicated to the proof of Lemma D.1. Fix $R_1 > \log(B)$.

### D.2.2  Realism performance

**Claim D.2.** *Since $R_1 > \log(B)$, a simple bound yields,*

$$(p^{\mathcal{U}}_{\lfloor\!\lfloor 2^{R_1} \rfloor\!\rfloor})^{\otimes B}(M^{(1)}, ..., M^{(B)} \text{ 2 by 2 distinct}) \geq 1 - \frac{B^2}{\lfloor 2^{R_1} \rfloor}.$$

See Appendix K for a proof. From the definition (Section D.2.1) of $Q$, for any $\mathcal{E} \in \mathcal{B}(\mathbb{R})$,

$$Q\left(\left\{\delta_0(\{y(\mathcal{C}, M^{(k)})\}_{k\in[B]}) \in \mathcal{E}\right\}\right.$$
$$\left.\left|\left\{M^{(1)}, ..., M^{(B)} \text{ 2 by 2 distinct}\right\}\right)\right.$$
$$= p_X^{\otimes B}\left(\delta_0(\mathbf{X}^{(B)}) \in \mathcal{E}\right). \tag{25}$$

Therefore, we have

$$\mathbb{E}_Q[\delta(\{y(\mathcal{C}, M^{(k)})\}_{k\in[B]})]$$
$$= \sum_{\mathbf{m}^{(B)}} \mathbb{E}_Q[\mathbf{1}_{\mathbf{M}^{(B)}=\mathbf{m}^{(B)}}\delta(\{y(\mathcal{C}, m^{(k)})\}_{k\in[B]})]$$
$$= \sum_{\mathbf{m}^{(B)}} \mathbb{E}_Q[\mathbf{1}_{\mathbf{M}^{(B)}=\mathbf{m}^{(B)}}]\mathbb{E}_Q[\delta(\{y(\mathcal{C}, m^{(k)})\}_{k\in[B]})]$$
$$= \sum_{\{m^{(k)}\}_{k\in[B]} \text{ 2 by 2} \neq} (p_{[\lfloor 2^{R_1}\rfloor]}^{\mathcal{U}})^{\otimes B}(\mathbf{M}^{(B)}=\mathbf{m}^{(B)})\mathbb{E}_{p_X^{\otimes B}}[\delta(\mathbf{X}^{(B)})]$$
$$+ \sum_{\{m^{(k)}\}_{k\in[B]} \text{ not 2 by 2} \neq} (p_{[\lfloor 2^{R_1}\rfloor]}^{\mathcal{U}})^{\otimes B}(\mathbf{M}^{(B)}=\mathbf{m}^{(B)})\mathbb{E}_Q[\delta(\{y(\mathcal{C}, m^{(k)})\}_{k\in[B]})]$$
$$\leq \mathbb{E}_{p_X^{\otimes B}}[\delta(\mathbf{X}^{(B)})] + \max(\delta)(p_{[\lfloor 2^{R_1}\rfloor]}^{\mathcal{U}})^{\otimes B}(M^{(1)}, ..., M^{(B)} \text{ not 2 by 2} \neq)$$
$$\leq \mathbb{E}_{p_X^{\otimes B}}[\delta(\mathbf{X}^{(B)})] + \frac{B^2}{\lfloor 2^{R_1}\rfloor} \max_x B \log \frac{1}{p_X(x)}, \tag{26}$$

where (26) follows from Claim D.2 and (3).

**Claim D.3.** *For any distribution $p$ on a finite set, any $p$-critic $\delta$ satisfies*

$$\mathbb{E}_p[\delta(X)] \leq 0. \tag{27}$$

*Proof.* By setting $q : \mathbf{x} \mapsto p(x) \cdot 2^{\delta(x)}$, we can write

$$\forall x \in \mathcal{X} \text{ s.t. } p(x) > 0, \ \delta(x) = \log\left(\frac{q(x)}{p(x)}\right), \text{ with } 0 < \sum_{x\in\mathcal{X}} q(x) \leq 1. \tag{28}$$

We denote the latter sum by $q(\mathcal{X})$. Then, $q/q(\mathcal{X})$ is a probability distribution on $\mathcal{X}$, and we have

$$\mathbb{E}_p[\delta(X)] \leq \mathbb{E}_p\left[\log\left(\frac{q(X)/q(\mathcal{X})}{p(X)}\right)\mathbf{1}_{p(X)>0}\right] = -KL(p\|q/q(\mathcal{X})) \leq 0. \tag{29}$$

$\square$

This concludes the proof of Lemma D.1.

## D.3 Further properties of a uniformly sampled batch

**Proposition D.4.** *Consider a finite set $\mathcal{X}$ such that $|\mathcal{X}| \geq 2$, a distribution $p_X$ on $\mathcal{X}$ such that $\forall x \in \mathcal{X}, p_X(x) > 0$, a positive integer $B$, some $R > \log(B)/\log(|\mathcal{X}|)$, some $\Delta \in \mathbb{R}_+$, and a $p_X^{\otimes B}$-critic $\delta$. Consider any conditional transition kernel $p_{Y|X}$ from $\mathcal{X}$ to $\mathcal{X}$ satisfying*

$$p_Y \equiv p_X, \quad \mathbb{E}_p[d(X, Y)] \leq \Delta. \tag{30}$$

*Then, for any $\varepsilon \in (0, \Delta/2)$, and any $\gamma > 0$, there exists a family $\{y(m)\}_{m\in[\lfloor 2^{R_1}\rfloor]}$, denoted $\mathbf{c}$, of elements of $\mathcal{X}$, such that distribution*

$$Q_{M,Y,X}(m, y, x) := \frac{1}{\lfloor 2^{R_1}\rfloor} \cdot \left(\mathbf{1}_{y=y(m)}\right) \cdot p_{X|Y=y(m)}(x) \tag{31}$$

*satisfies*

$$\|Q_X - p_X\|_{TV} \leq \frac{3\Delta}{\varepsilon}[p(\mathcal{A}_{R,\gamma}) + 2^{-\gamma \log(|\mathcal{X}|)/2}] \tag{32}$$

$$\mathbb{E}_{Q^{\otimes B}}[d(\mathbf{X}^{(B)}, \mathbf{Y}^{(B)})] \leq \Delta + \varepsilon \tag{33}$$

$$\mathbb{E}_{Q^{\otimes B}}[\delta(\mathbf{Y}^{(B)})] \leq \frac{3\Delta}{\varepsilon} \cdot \frac{B^2}{\lfloor 2^{R_1} \rfloor} \max_x B \log \frac{1}{p_X(x)}, \tag{34}$$

*where* $R_1 = R \log|X|$, *and*

$$\mathcal{A}_{R,\gamma} := \left\{ (x,y) \in \mathcal{X}^2 \mid \log\left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\right) - \log(\lfloor 2^{R_1} \rfloor) > -\gamma \log(|\mathcal{X}|) \right\}. \tag{35}$$

*Proof.* Fix some $R > \log(B)/\log(|\mathcal{X}|)$, some $\Delta > 0$, some $\varepsilon \in (0, \Delta/2)$, some $\gamma > 0$, and a conditional transition kernel $p_{Y|X}$ from $\mathcal{X}$ to $\mathcal{X}$ satisfying

$$p_Y \equiv p_X, \quad \mathbb{E}_p[d(X,Y)] \leq \Delta. \tag{36}$$

Define $R_1 = R \log|\mathcal{X}|$. We apply Lemma D.1, and use the notation therein. Then, from Markov's inequality, we have

$$Q_\mathcal{C}\left(\mathbb{E}_Q[\delta(\mathbf{Y}^{(B)})|\mathcal{C}] \geq \frac{3\Delta}{\varepsilon} \frac{B^2}{\lfloor 2^{R_1} \rfloor} \max_x B \log \frac{1}{p_X(x)}\right) \leq \frac{\varepsilon}{3\Delta}. \tag{37}$$

We extend distribution $Q$ as follows.

$$Q_{\mathcal{C}, \mathbf{M}^{(B)}, \mathbf{Y}^{(B)}, \mathbf{X}^{(B)}}\left(\{y(m')\}_{m' \in [\lfloor 2^{R_1} \rfloor]}, \mathbf{m}^{(B)}, \mathbf{y}^{(B)}, \mathbf{x}^{(B)}\right) :=$$

$$Q_{\mathcal{C}, \mathbf{M}^{(B)}, \mathbf{Y}^{(B)}}\left(\{y(m')\}_{m' \in [\lfloor 2^{R_1} \rfloor]}, \mathbf{m}^{(B)}, \mathbf{y}^{(B)}\right) \cdot \prod_{k=1}^B p_{X|Y=y(m^{(k)})}(x^{(k)}). \tag{38}$$

Distribution $Q_{\mathcal{C}, M^{(1)}, Y^{(1)}, X^{(1)}}$ corresponds to the setting of Theorem VII.1 of [7], known as the soft covering lemma. Since $p_Y \equiv p_X$, the latter lemma yields that for any $\tau \in \mathbb{R}$,

$$\mathbb{E}_\mathcal{C}\left[\|Q_{X^{(1)}|\mathcal{C}} - p_X\|_{TV}\right] \leq p(\mathcal{A}_\tau) + 2^{\tau/2}, \tag{39}$$

where

$$\mathcal{A}_\tau := \{(x,y) \mid \log(p_{Y|X=x}(y)/p_X(y)) - \log(\lfloor 2^{R_1} \rfloor) > \tau\}. \tag{40}$$

We choose $\tau = -\gamma \log|\mathcal{X}|$. Then, $\mathcal{A}_\tau = \mathcal{A}_{R,\gamma}$, with the notation of Proposition D.4. Hence, from (39) and Markov's inequality, we have

$$Q_\mathcal{C}\left(\|Q_{X^{(1)}|\mathcal{C}} - p_X\|_{TV} \geq \frac{3\Delta}{\varepsilon}[p(\mathcal{A}_{R,\gamma}) + 2^{-\gamma \log(|\mathcal{X}|)/2}]\right) \leq \frac{\varepsilon}{3\Delta}. \tag{41}$$

By construction, we have $Q_{\mathbf{Y}^{(B)}, \mathbf{X}^{(B)}} \equiv p_{Y,X}^{\otimes B}$. Therefore, from (36), and the additivity of $d$, we have

$$\mathbb{E}_Q[d(\mathbf{X}^{(B)}, \mathbf{Y}^{(B)})] \leq \Delta. \tag{42}$$

Therefore, from Markov's inequality,

$$Q_\mathcal{C}\left(\mathbb{E}_Q[d(\mathbf{X}^{(B)}, \mathbf{Y}^{(B)})|\mathcal{C}] \geq \Delta + \varepsilon\right) \leq \frac{\Delta}{\Delta + \varepsilon} = 1 - \frac{\varepsilon}{\Delta} \cdot \frac{1}{1 + \varepsilon/\Delta} < 1 - \frac{2\varepsilon}{3\Delta}, \tag{43}$$

where we have used the fact that $\varepsilon \in (0, \Delta/2)$. From a union bound and (37), (41), and (43) there exists a realization $c_*$ of $\mathcal{C}$ such that none of the corresponding events hold. Since, by construction,

$$Q_{\mathbf{M}^{(B)}, \mathbf{Y}^{(B)}, \mathbf{X}^{(B)}|\mathcal{C}=c*} \equiv Q_{M^{(1)}, Y^{(1)}, X^{(1)}|\mathcal{C}=c*}^{\otimes B},$$

this concludes the proof of Proposition D.4. $\qquad \square$

## D.4 Proof of Theorem C.1

Fix some $R > \log(B)/\log(|\mathcal{X}|)$, some $\Delta > 0$, some $\varepsilon \in (0, \Delta/2)$, some $\gamma > 0$, and a conditional transition kernel $p_{Y|X}$ from $\mathcal{X}$ to $\mathcal{X}$ satisfying

$$p_Y \equiv p_X, \quad \mathbb{E}_p[d(X,Y)] \leq \Delta. \tag{44}$$

Define $R_1 = R \log|\mathcal{X}|$. Then, we can apply Proposition D.4. We use the notation from the latter.

### D.4.1 Compression scheme achieving close-to-uniform sampling

We define the following distribution $P_{X,Y,M}$, which differs from $Q$ in having the correct marginal for $X$:

$$P_{X,M,Y} := p_X \cdot Q_{M,Y|X}. \tag{45}$$

Therefore, from (31), distribution $P$ satisfies Markov chain $X-M-Y$. Hence, it defines a $(R_1, 0)$ code. From Lemma J.2 (Appendix J), comparing $P$ with $Q$ reduces to comparing marginals, i.e. to (32):

$$\begin{aligned} \left\| P_{M,X,Y} - Q_{M,X,Y} \right\|_{TV} &= \left\| P_X - Q_X \right\|_{TV} \\ &= \left\| p_X - Q_X \right\|_{TV} \le \frac{3\Delta}{\varepsilon} [p(\mathcal{A}_{R,\gamma}) + 2^{-\gamma \log(|\mathcal{X}|)/2}]. \end{aligned} \tag{46}$$

Since $d$ is additive, we have

$$\mathbb{E}_{(P)^{\otimes B}}[d(\mathbf{X}^{(B)}, \mathbf{Y}^{(B)})] = \mathbb{E}_P[d(X,Y)] \text{ and}$$

$$\mathbb{E}_{(Q)^{\otimes B}}[d(\mathbf{X}^{(B)}, \mathbf{Y}^{(B)})] = \mathbb{E}_Q[d(X,Y)].$$

Since $d$ is bounded, then we can apply Lemma J.3 (Appendix J). Then, from (46), and Lemma J.1 with $W = (X,Y)$, we have

$$\begin{aligned} \mathbb{E}_{P^{\otimes B}}[d(\mathbf{X}^{(B)}, \mathbf{Y}^{(B)})] &\le \mathbb{E}_{Q^{\otimes B}}[d(\mathbf{X}^{(B)}, \mathbf{Y}^{(B)})] + \frac{6\Delta}{\varepsilon} \max(d)[p(\mathcal{A}_{R,\gamma}) + 2^{-\gamma \log(|\mathcal{X}|)/2}] \\ &\le \Delta + \varepsilon + \frac{6\Delta}{\varepsilon} \max(d)[p(\mathcal{A}_{R,\gamma}) + 2^{-\gamma \log(|\mathcal{X}|)/2}], \end{aligned} \tag{47}$$

where the last inequality follows from (33). Moving to the realism performance, we have the following property of the TVD - see Appendix J:

**Claim D.5.** *Given any two distributions $P$ and $Q$ on the same finite alphabet, we have, for any $B \in \mathbb{N}$,*

$$\left\| P^{\otimes B} - Q^{\otimes B} \right\|_{TV} \le B \left\| P - Q \right\|_{TV}.$$

From Lemma J.3, Claim D.5, (46), and Lemma J.1 with $W = \mathbf{Y}^{(B)}$, we have,

$$\mathbb{E}_{P^{\otimes B}}[\delta(\mathbf{Y}^{(B)})] \le \mathbb{E}_{Q^{\otimes B}}[\delta(\mathbf{Y}^{(B)})] + \frac{6B\Delta}{\varepsilon} \max(\delta)[p(\mathcal{A}_{R,\gamma}) + 2^{-\gamma \log(|\mathcal{X}|)/2}]$$

$$\le \frac{3\Delta}{\varepsilon} \cdot \frac{B^2}{\lfloor 2^{R_1} \rfloor} \max_x B \log \frac{1}{p_X(x)} + \frac{6B\Delta}{\varepsilon} [p(\mathcal{A}_{R,\gamma}) + 2^{-\gamma \log(|\mathcal{X}|)/2}] \cdot \max_x B \log \frac{1}{p_X(x)}. \tag{48}$$

This concludes the proof.

## E  Achievability of Theorem 4.1

Consider some $\Delta \in \mathbb{R}_+$ such that $R^{(1)}(\Delta, 0) < H_p(X)$, and a sequence $\{B_n\}_{n \ge 1}$ of positive integers such that

$$\log(B_n)/n \xrightarrow[n \to \infty]{} 0. \tag{49}$$

Fix $R \in (R^{(1)}(\Delta, 0), H_p(X))$, $\varepsilon \in (0, R - R^{(1)}(\Delta, 0))$, and $\gamma \in (0, \varepsilon/\log(|\mathcal{X}|))$. Then, there exists $p_{Y|X}$ such that

$$p_Y \equiv p_X, \ \mathbb{E}_p[d(X,Y)] \le \Delta, \ R \ge I_p(X;Y) + \varepsilon. \tag{50}$$

We use the powerful result of Theorem B.4 regarding the existence of a so-called *universal critic*. From Definition 2.1, for every $n \in \mathbb{N}$, the restriction of $\delta_0$ to $\mathcal{X}^{nB_n}$ is a $p_X^{\otimes nB_n}$-critic. Moreover, from (49), for large enough $n$, we have $nR > \log(B_n)$. Then, for large enough $n$, we can apply Theorem C.1 for set $\mathcal{X}^n$, distribution $p_X^{\otimes n}$, transition kernel $\prod_{t=1}^n p_Y|X$, batch size $B_n$, critic $\delta_0$, rate

14

$nR/\log(|\mathcal{X}^n|)$, and constants $\Delta, \varepsilon, \gamma$. This gives that, for every $n$ large enough, there is a $(n, R, 0)$ code, inducing a distribution $P^{(n)}$ such that

$$\mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[d(\mathbf{X}^{(n,B_n)}, \mathbf{Y}^{(n,B_n)})\right] \leq \Delta + \varepsilon + \frac{6\Delta}{\varepsilon} \max(d)[p(\mathcal{A}_{R,\gamma}^{(n)}) + 2^{-\gamma n \log(|\mathcal{X}|)/2}], \quad (51)$$

$$\mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[\delta_0(\mathbf{Y}^{(n,B_n)})\right] \leq$$

$$\frac{3\Delta}{\varepsilon}\left[\frac{B_n^2}{\lfloor 2^{nR}\rfloor}\max_x nB_n \log \frac{1}{p_X(x)} + 2B_n[p(\mathcal{A}_{R,\gamma}^{(n)}) + 2^{-\gamma n \log(|\mathcal{X}|)/2}] \cdot \max_x nB_n \log \frac{1}{p_X(x)}\right], \quad (52)$$

where

$$\mathcal{A}_{R,\gamma}^{(n)} := \left\{(x_{1:n}, y_{1:n}) \in (\mathcal{X}^n)^2 \mid \sum_{t=1}^n \log\left(\frac{p_{X,Y}(x_t, y_t)}{p_X(x_t)p_Y(y_t)}\right) - \log(\lfloor 2^{nR}\rfloor) > -\gamma n \log(|\mathcal{X}|)\right\}, \quad (53)$$

with the convention $0/0 := 1$. From (50), $\log(\lfloor 2^{nR}\rfloor)/n - \gamma \log(|\mathcal{X}|) > I_p(X; Y)$ for large enough $n$. Then, since $\mathcal{X}$ is finite, we have, from Hoeffding's inequality,

$$p(\mathcal{A}_{R,\gamma}^{(n)}) = O(e^{-\kappa n}), \quad (54)$$

for some $\kappa > 0$. Hence, from (49), (51), (52), and Theorem B.4, we have

$$\limsup_{n\to\infty} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[d(\mathbf{X}^{(n,B_n)}, \mathbf{Y}^{(n,B_n)})\right] \leq \Delta + \varepsilon, \quad (55)$$

and for any lower semi-computable $p_X^{\otimes *}$-critic $\delta$,

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[\delta(\mathbf{Y}^{(n,B_n)})\right] < \infty. \quad (56)$$

From the proof of Theorem C.1, we know that $P^{(n)}$ has a deterministic decoder. Hence, it only remains to derandomize the encoder of $P^{(n)}$. We denote its decoder by $m \mapsto y_{1:n}(m)$. The following claim is a slight modification of Proposition 4 in [10]. We provide details in Section E.1.

**Claim E.1.** *There exists a sequence of deterministic maps*

$$f^{(n)} : \mathcal{X}^n \to [2^{nR}], \quad \text{such that}$$

$$\left\|\hat{\tilde{P}}_{\mathcal{X}^2}^{(n)}[X^n, y_{1:n}(M)] - \hat{P}_{\mathcal{X}^2}^{(n)}[X^n, y_{1:n}(M)]\right\|_{TV} \xrightarrow[n\to\infty]{} 0,$$

$$\liminf_{n\to\infty} \frac{-1}{n}\log\left\|\tilde{P}_M^{(n)} - P_M^{(n)}\right\|_{TV} > 0, \quad \text{where} \quad (57)$$

$$\tilde{P}_{X^n, M}^{(n)} := p_X^{\otimes n} \cdot \mathbf{1}_{M=f^{(n)}(X^n)}.$$

Then, from (49) and Claim D.5, we have

$$\liminf_{n\to\infty} \frac{-1}{n}\log\left\|(\tilde{P}^{(n)})_M^{\otimes B_n} - (P^{(n)})_M^{\otimes B_n}\right\|_{TV} > 0. \quad (58)$$

Thus, from Lemma J.3 and (3), we have

$$\left|\mathbb{E}_{(\tilde{P}^{(n)})^{\otimes B_n}}\left[\delta(\mathbf{Y}^{(n,B_n)})\right] - \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[\delta(\mathbf{Y}^{(n,B_n)})\right]\right| \xrightarrow[n\to\infty]{} 0. \quad (59)$$

Moreover, since $d$ is bounded, then from Lemma J.3, we obtain

$$\left|\mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[d(\mathbf{X}^{(n,B_n)}, \mathbf{Y}^{(n,B_n)})\right] - \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[d(\mathbf{X}^{(n,B_n)}, \mathbf{Y}^{(n,B_n)})\right]\right| \xrightarrow[n\to\infty]{} 0. \quad (60)$$

Since this analysis is valid for any $\varepsilon \in (0, R - R^{(1)}(\Delta, 0))$, then tuple $(R, \{B_n\}_{n\geq 1}, \Delta)$ is asymptotically achievable with algorithmic realism with fully deterministic codes. This being true for any $R \in (R^{(1)}(\Delta, 0), H_p(X))$, we have

$$R(\Delta) \leq R_*(\Delta) \leq R^{(1)}(\Delta, 0),$$

as desired.

## E.1 Encoder derandomization

We show that Claim E.1 follows from Proposition 4 in [10], and its proof. We can apply that result directly, since $R < H_p(X)$ and $\mathcal{X}$ is finite. This would give all properties in Claim E.1, except for the exponential decay in (57). To obtain the latter, it is sufficient to adapt the proof of Proposition 4 in [10], by replacing the use of the law of large numbers with the use of Hoeffding's inequality, and using Theorem VII.1 of [7] with $\tau = -n\gamma$, for small enough $\gamma$.

# F Converse of Theorem 4.1

From standard information-theoretic arguments, we have the following result - see Appendix H for a proof.

**Lemma F.1.** *Consider a triplet $(R, R_c, \Delta)$ and a sequence of codes, the $n$-th being $(n, R, R_c)$, inducing a sequence $\{P^{(n)}_{X_{1:n}, J, M, Y_{1:n}}\}_{n \geq 1}$ of distributions such that*

$$\limsup_{n \to \infty} \mathbb{E}_{(P^{(n)})^{\otimes b_n}} \left[ d(\mathbf{X}^{(n, b_n)}, \mathbf{Y}^{(n, b_n)}) \right] \leq \Delta, \tag{61}$$

*for some sequence $\{b_n\}_{n \geq 1}$ of positive integers. For every $n \geq 1$, let $T^{(n)}$ denote a uniform variable on $[nb_n]$ independent from all other random variables. Then, there exists a conditional distribution $p_{Y|X}$ and an increasing sequence $\{n_i\}_{i \geq 1}$ of positive integers such that*

$$(P^{(n_i)})^{\otimes b_{n_i}}_{X_{T^{(n_i)}}, Y_{T^{(n_i)}}} \xrightarrow[i \to \infty]{} p_{X,Y} \tag{62}$$

$$\Delta \geq \mathbb{E}_p[d(X, Y)] \tag{63}$$

$$R \geq I_p(X; Y), \tag{64}$$

*where $p_{X,Y}$ refers to $p_X \cdot p_{Y|X}$.*

## F.1 Converse proof

Consider some $\Delta \in \mathbb{R}_+$ such that $R^{(1)}(\Delta, 0) < H_p(X)$, and a sequence $\{B_n\}_{n \geq 1}$ of positive integers such that

$$\log(B_n)/n \xrightarrow[n \to \infty]{} 0. \tag{65}$$

We know that $R_*(\Delta) \geq R(\Delta)$, and prove that $R(\Delta) \geq R^{(1)}(\Delta, 0)$. Consider a couple $(R, \Delta) \in \mathbb{R}_+^2$, and some $R_c \in \mathbb{R}_{\geq 0}$ such that $(R, R_c, \{B_n\}_{n \geq 1}, \Delta)$ is asymptotically achievable with algorithmic realism. Fix $\varepsilon > 0$. Then, there exists a sequence of codes, the $n$-th being $(n, R, R_c)$, inducing a sequence $\{P^{(n)}_{X_{1:n}, J, M, Y_{1:n}}\}_n$ of distributions such that

$$\limsup_{n \to \infty} \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \left[ d(\mathbf{X}^{(n, B_n)}, \mathbf{Y}^{(n, B_n)}) \right] \leq \Delta + \varepsilon, \tag{66}$$

and for any lower semi-computable $p_X^{\otimes *}$-critic $\delta$,

$$\sup_{n \in \mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \left[ \delta(\mathbf{Y}^{(n, B_n)}) \right] < \infty. \tag{67}$$

Then, Lemma F.1 applies, with $b_n = B_n$, for all $n$, with $R + \varepsilon$ instead of $R$, and $\Delta + \varepsilon$ instead of $\Delta$. Then, there exists a conditional distribution $p_{Y|X}$ and an increasing sequence $\{n_i\}_{i \geq 1}$ of positive integers such that

$$(P^{(n_i)})^{\otimes b_{n_i}}_{X_{T^{(n_i)}}, Y_{T^{(n_i)}}} \xrightarrow[i \to \infty]{} p_{X,Y} \tag{68}$$

$$\Delta + \varepsilon \geq \mathbb{E}_p[d(X, Y)] \tag{69}$$

$$R + \varepsilon \geq I_p(X; Y), \tag{70}$$

where for any $n \in \mathbb{N}$, variable $T^{(n)}$ is uniformly distributed on $[nB_n]$, and independent from all other random variables. We prove that $p_Y \equiv p_X$. Fix $e_0 \in \mathcal{X}$. Consider the computable $p_X^{\otimes *}$-critic $\delta$ from Claim I.1, with $q$ therein taken to be $p_X$. Then, from (67),

$$\sup_{n \in \mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \left[ \delta(\mathbf{Y}^{(n, B_n)}) - 2 \log(\delta(\mathbf{Y}^{(n, B_n)}) + 3) \right] < \infty. \tag{71}$$

Thus,

$$\sup_{n \in \mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \left[ \delta(\mathbf{Y}^{(n, B_n)}) \right] < \infty, \quad \text{and} \quad \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \left[ \delta(\mathbf{Y}^{(n, B_n)}) - \frac{1}{2} \log(nB_n) \right] \xrightarrow[n \to \infty]{} -\infty.$$

Thus, the frequency of $e_0$ in a batch of reconstructions converges in $L_1$ norm to $p_X(e_0)$. Hence, the expected frequencies converge to $p_X(e_0)$. This rewrites as

$$(P^{(n)})^{\otimes B_n}_{Y_{T^{(n)}}}(e_0) \to p_X(e_0). \tag{72}$$

This is true for any $e_0$ in $\mathcal{X}$. Thus, from (68), $p_Y \equiv p_X$. Hence, from (69) and (70), we have

$$R + \varepsilon \geq R^{(1)}(\Delta + \varepsilon, 0). \tag{73}$$

This being true for any $\varepsilon > 0$, and since $R^{(1)}(\cdot, 0)$ is convex -thus continuous- on $(0, \infty)$, we have

$$R \geq R^{(1)}(\Delta, 0). \tag{74}$$

This being true for any $R \in \mathbb{R}_+$ such that there exists $R_c \in \mathbb{R}_{\geq 0}$ such that $(R, R_c, \{B_n\}_{n \geq 1}, \Delta)$ is asymptotically achievable with algorithmic realism, we have

$$R(\Delta) \geq R^{(1)}(\Delta, 0), \tag{75}$$

as desired.

# G   Proof of Theorem 4.3

Consider an increasing sequence $\{B_n\}_{n \geq 1}$ of positive integers such that

$$\frac{B_n}{|\mathcal{X}|^n} \to \infty, \tag{76}$$

some $R_c \in \mathbb{R}_{\geq 0}$, and some $(R, \Delta) \in (\mathbb{R}_+)^2$ such that tuple $(R, R_c, \Delta)$ is asymptotically achievable with near-perfect realism. From Theorem 1 in [24], $(R, R_c, \Delta)$ achievable with *perfect realism*, i.e. satisfying the properties in Definition 4.2, with (12) replaced with

$$\exists N \in \mathbb{N}, \forall n \geq N, \ P_{Y_{1:n}}^{(n)} \equiv p_X^{\otimes n}. \tag{77}$$

Fix $\varepsilon > 0$, and a corresponding sequence of $(n, R + \varepsilon, R_c)$ codes. Denote by $P^{(n)}$ the distribution induces by the $n$-th code. Then, there exists an integer $N_\varepsilon$ such that

$$\limsup_{n \to \infty} \mathbb{E}_{P^{(n)}}\big[d(X_{1:n}, Y_{1:n})\big] \leq \Delta + \varepsilon, \tag{78}$$

$$\forall n \geq N_\varepsilon, \ (P_{Y_{1:n}}^{(n)})^{\otimes B_n} \equiv p_X^{\otimes n B_n}. \tag{79}$$

From (78), (79), Claim D.3, and the additivity of the distortion measure $d$, we have

$$\limsup_{n \to \infty} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\big[d(\mathbf{X}^{(n, B_n)}, \mathbf{Y}^{(n, B_n)})\big] \leq \Delta + \varepsilon, \tag{80}$$

and for any lower semi-computable $p_X^{\otimes *}$-critic $\delta$,

$$\sup_{n \in \mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\big[\delta(\mathbf{Y}^{(n, B_n)})\big] < \infty. \tag{81}$$

Since this analysis is valid for every $\varepsilon > 0$, then $(R, R_c, \{B_n\}_{n \geq 1}, \Delta)$ is asymptotically achievable with algorithmic realism. Moving to the converse, consider a computable increasing sequence $\{B_n\}_{n \geq 1}$ of positive integers such that

$$\frac{B_n}{|\mathcal{X}|^n} \to \infty, \tag{82}$$

some $R_c \in \mathbb{R}_{\geq 0}$, and some $(R, \Delta) \in (\mathbb{R}_+)^2$ such that tuple $(R, R_c, \Delta)$ is asymptotically achievable with algorithmic realism. Fix $\varepsilon > 0$. Then, there exists a sequence of codes, the $n$-th being $(n, R + \varepsilon, R_c)$, such that the sequence $\{P^{(n)}\}_n$ of distributions induced by the codes satisfies

$$\limsup_{n \to \infty} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\big[d(\mathbf{X}^{(n, B_n)}, \mathbf{Y}^{(n, B_n)})\big] \leq \Delta + \varepsilon, \ \text{ and} \tag{83}$$

and for any lower semi-computable $p_X^{\otimes *}$-critic $\delta$,

$$\sup_{n \in \mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\big[\delta(\mathbf{Y}^{(n, B_n)})\big] < \infty. \tag{84}$$

**Lemma G.1.** *[4] There exists a positive integer $\lambda$ such that for any $k \in \mathbb{N}$, any distribution $q$ on some finite set $\mathcal{W}$ of size $k$, any $\varepsilon, \eta > 0$, and any integer $b$ satisfying*

$$b \geq \lambda \cdot \frac{k + \log(1/\eta)}{\varepsilon^2}, \tag{85}$$

*we have*

$$q^{\otimes b}\Big(\big\|\mathbb{P}_{\mathcal{W}}^{emp}[W^b] - q\big\|_{TV} \geq \varepsilon\Big) \leq \eta. \tag{86}$$

17

For every $n \in \mathbb{N}$, define

$$C_n := \left\lceil \left( \frac{B_n}{|\mathcal{X}|^n} \right)^{\frac{1}{3}} \right\rceil. \tag{87}$$

Since $\mathcal{X}$ is finite, $\{C_n\}_{n \geq 1}$ is a computable sequence of positive integers. Moreover, from (82), we have

$$C_n \underset{n \to \infty}{\longrightarrow} \infty. \tag{88}$$

Choosing, for every $n \in \mathbb{N}$, $\eta = 1/3$ and $\varepsilon = 1/C_n$, then from Lemma G.1 and (88) we have, for large enough $n$,

$$(P^{(n)})^{\otimes B_n} \left( \left\| \mathbb{P}^{\mathrm{emp}}_{\mathcal{X}^n}[\mathbf{Y}^{(n,B_n)}] - P^{(n)}_{Y_{1:n}} \right\|_{TV} \geq \frac{1}{C_n} \right) \leq \frac{1}{3}. \tag{89}$$

Consider the computable sequence of positive integers defined by

$$\forall n \in \mathbb{N}, \; A_n := \left\lceil \left( \frac{B_n}{|\mathcal{X}|^n} \right)^{\frac{4}{9}} \right\rceil. \tag{90}$$

Since $\{B_n\}_{n \geq 1}$ is increasing, then for any $t \in \mathbb{N}$, there exists a unique integer $n \in \mathbb{N}_{\geq 0}$ such that

$$t \in [nB_n, (n+1)B_{n+1}),$$

with the definition $B_0 := 0$. We define $\delta : \cup_{t \in \mathbb{N}} \mathcal{X}^t \to \mathbb{N}_{\geq 0}$ as follows. For any integer $t \in [1, B_1)$, and any $x_{1:t} \in \mathcal{X}^t$, let $\delta(x) := 0$. For any $n \in \mathbb{N}$, any $t \in [nB_n, (n+1)B_{n+1})$, and any $x_{1:t} \in \mathcal{X}^t$, let

$$\delta(x_{1:t}) := \left\lceil A_n \left\| \mathbb{P}^{\mathrm{emp}}_{\mathcal{X}^n}[x_{1:nB_n}] - p_X^{\otimes n} \right\|_{TV} \right\rceil. \tag{91}$$

**Claim G.2.** *From Lemma G.1 and* (88)*, there exists a positive integer $L$ such that $\delta - 2\log(\delta+3) - L$ is a lower semi-computable $p_X^{\otimes *}$-critic.*

We provide a proof in Appendix I.2. Then, we can apply (84) to critic $\delta - 2\log(\delta + 3) - L$, and get,

$$\sup_{n \in \mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \left[ \delta(\mathbf{Y}^{(n,B_n)}) - 2\log(\delta(\mathbf{Y}^{(n,B_n)}) + 3) - L \right] < \infty. \tag{92}$$

Thus,

$$\sup_{n \in \mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \left[ \delta(\mathbf{Y}^{(n,B_n)}) \right] < \infty, \; \text{and} \; (P^{(n)})^{\otimes B_n} \left( \delta(\mathbf{Y}^{(n,B_n)}) \geq C_n \right) \underset{n \to \infty}{\longrightarrow} 0,$$

because $\{C_n\}_{n \geq 1}$ tends to infinity. Combining this with (89) through a union bound, we obtain, from the triangle inequality for the TVD,

$$(P^{(n)})^{\otimes B_n} \left( \left\| P^{(n)}_{Y_{1:n}} - p_X^{\otimes n} \right\|_{TV} \leq \frac{C_n}{A_n} + \frac{1}{C_n} \right) > 0,$$

for large enough $n$. The above event does not depend on the random batch, hence the corresponding inequality is true, for large enough $n$. Since $\{C_n\}_{n \geq 1}$ tends to infinity and since from (82), (87), and (90), we have $C_n/A_n \to 0$, then we obtain

$$\left\| P^{(n)}_{Y_{1:n}} - p_X^{\otimes n} \right\|_{TV} \underset{n \to \infty}{\longrightarrow} 0. \tag{93}$$

Hence, from (83) and the additivity of $d$, we have that $(R, R_c, \Delta)$ is asymptotically achievable with near-perfect realism. This concludes the proof.

## H   Standard converse arguments

Here, we provide a proof of Lemma F.1 (Appendix F). The sequence of distributions $(P^{(n)})^{\otimes b_n}_{X_T, Y_T}$ can be seen as a bounded sequence in $\mathbb{R}^{2^{2s}}$, thus it admits a converging subsequence:

$$(P^{(n_i)})^{\otimes b_n}_{X_T, Y_T} \underset{i \to \infty}{\longrightarrow} p_{X,Y}. \tag{94}$$

Since $d$ is bounded, we have

$$\mathbb{E}_{(P^{(n_i)})^{\otimes b_n}}[d(X_T, Y_T)] \xrightarrow[i \to \infty]{} \mathbb{E}_p[d(X, Y)]. \tag{95}$$

Since $d$ is additive, we have, for any $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}_{(P^{(n)})^{\otimes b_n}} & \left[ d(\mathbf{X}^{(n,b_n)}, \mathbf{Y}^{(n,b_n)}) \right] \\
&= \mathbb{E}_{(P^{(n)})^{\otimes b_n}} \left[ d(X_T, Y_T) \right].
\end{aligned} \tag{96}$$

From (61), (95) and (96), we have $\Delta \geq \mathbb{E}_p[d(X, Y)]$. Secondly, distribution $P^{(n)}$ satisfies

$$\begin{aligned}
nb_n R &\geq H(\{m^{(k)}\}_{k \in [b_n]} | \{J^{(k)}\}_{k \in [b_n]}) \\
&\geq I(\{m^{(k)}\}_{k \in [b_n]}; \mathbf{X}^{(n,b_n)} | \{J^{(k)}\}_{k \in [b_n]}) \\
&= I(\{m^{(k)}\}_{k \in [b_n]}, \{J^{(k)}\}_{k \in [b_n]}; \mathbf{X}^{(n,b_n)}) \\
&\geq I(\mathbf{Y}^{(n,b_n)}; \mathbf{X}^{(n,b_n)}) \\
&\geq \sum_{k=1}^{b_n} \sum_{t=1}^{n} I(Y_t^{(k)}; X_t^{(k)}) \\
&= nb_n I(Y_T; X_T | T) \\
&= nb_n I(T, Y_T; X_T) \\
&\geq nb_n I(Y_T; X_T).
\end{aligned}$$

Therefore, from (94), and by continuity of mutual information on the set of distributions on $(\{0,1\}^s)^2$, we have $R \geq I_p(X; Y)$.

# I   Frequency critics

## I.1   Critic involving the frequency of a specific pattern

The following claim, and its proof, are inspired from Lemma 4.3.5 and Exercise 2.4.1 in [14].

**Claim I.1.** *Consider a finite set $\mathcal{X}$, identified with a subset of $\{0,1\}^s$. Let $q$ be a distribution on $\mathcal{X}$ such that $\forall x \in \mathcal{X}, q(x) > 0$. Let $e_0$ be any string in $\mathcal{X}$, considered as a pattern of interest. For any $n \in \mathbb{N}$ and any $x_{1:n} \in \mathcal{X}^n$, let $S(x_{1:n})$ denote the number of occurrences of $e_0$ in $x_{1:n}$. Define map $\delta: \cup_{n \in \mathbb{N}} \to \mathbb{N}_{\geq 0}$ by*

$$\forall n \in \mathbb{N}, \forall x_{1:n} \in \mathcal{X}^n, \quad x_{1:n} \mapsto \left\lceil \log \left\lceil |S(x_{1:n}) - q(e_0)n| / \sqrt{n} \right\rceil \right\rceil. \tag{97}$$

*Then, $\delta - 2\log(\delta + 3)$ is a computable $q^{\otimes *}$-critic.*

*Proof.* From Lemma B.2, $\delta$ is lower semi-computable. Since $\delta - 2\log(\delta + 3) = \log(2^\delta / (\delta + 3)^2)$, then by Lemma B.2, $\delta - 2\log(\delta + 3)$ is lower semi-computable. For any $(n, C) \in \mathbb{N}^2$, and any $x_{1:n} \in \mathcal{X}^n$, we have:

$$\begin{aligned}
\{\delta(x_{1:n}) \geq C\} &= \left\{ \left\lceil \log \left\lceil |S(x_{1:n}) - nq(e_0)| / \sqrt{n} \right\rceil \right\rceil \geq C \right\} \\
&= \left\{ \log \left\lceil |S(x_{1:n}) - nq(e_0)| / \sqrt{n} \right\rceil > C - 1 \right\} \\
&= \left\{ \left\lceil |S(x_{1:n}) - nq(e_0)| / \sqrt{n} \right\rceil > 2^{C-1} \right\} \\
&= \left\{ |S(x_{1:n}) - nq(e_0)| / \sqrt{n} > 2^{C-1} \right\}.
\end{aligned}$$

From this and Chebyshev inequality, we obtain:

$$\begin{aligned}
q^{\otimes n}(\delta(X_{1:n}) \geq C) &\leq \mathbb{E}_{q^{\otimes n}} \left[ (S(X_{1:n}) - nq(e_0))^2 / n \right] / 4^{C-1} \\
&= (q(e_0) - q(e_0)^2) / 4^{C-1} \\
&\leq 4^{-C} \\
&\leq 2^{-C},
\end{aligned} \tag{98}$$

19

where (98) comes from the fact that $S(X_{1:n})$ follows a binomial distribution $B(n, q(e_0))$. Thus,

$$\mathbb{E}_{q^{\otimes n}}[\mathbf{1}_{\delta(X_{1:n})=C}] \leq 2^{-C}. \tag{99}$$

Therefore,

$$\mathbb{E}_{q^{\otimes n}}[\mathbf{1}_{\delta(X_{1:n})=C} \cdot 2^{\delta(X_{1:n})-2\log(\delta(X_{1:n})+3)}] \leq \frac{1}{(C+3)^2}. \tag{100}$$

This also holds for $C = 0$. Summing over $C \in \mathbb{N}_{\geq 0}$ gives, for any $n \in \mathbb{N}$,

$$\sum_{x_{1:n} \in \mathcal{X}^n} q^{\otimes n}(x_{1:n}) \cdot 2^{\delta(x_{1:n})-2\log(\delta(x_{1:n})+1)-1} \leq 1. \tag{101}$$

Hence, we have that $\delta - 2\log(\delta + 3)$ is a lower semi-computable $q^{\otimes *}$-critic. □

## I.2 Critic involving an empirical distribution

We provide a proof of Claim G.2.

**Claim I.2.** *The map* $f : \cup_{t \in \mathbb{N}} \mathcal{X}^t \to \mathbb{R}$ *defined by* $\forall t \in [1, B_1] \cap \mathbb{N}, \forall x_{1:t} \in \mathcal{X}^t, f(x_{1:t}) := 0$, *and*

$$\forall n \in \mathbb{N}, \forall t \in [nB_n, (n+1)B_{n+1}) \cap \mathbb{N}, \forall x_{1:t} \in \mathcal{X}^t, \ f(x_{1:t}) := \left\|\mathbb{P}^{emp}_{\mathcal{X}^n}[x_{1:nB_n}] - p^{\otimes n}\right\|_{TV} \ (102)$$

*is computable.*

*Proof.* Since there exists $s \in \mathbb{N}$ such that $\mathcal{X} \subseteq \{0, 1\}^s$, then, given some $x \in \cup_{t \in \mathbb{N}} \mathcal{X}^t \to \mathbb{R}$, one can compute the unique corresponding $t$ via a Turing machine. Moreover, since $\{B_n\}_{n \geq 1}$ is computable, one can further compute the unique $n$ such that $t \in [nB_n, (n+1)B_{n+1})$ via a Turing machine, as well as the empirical probability appearing in (102). For any $k \in \mathbb{N}$, and any $x_0 \in \mathcal{X}$, one can call the rational-valued computable upper and lower approximations of $p$ at point $(x_0, k)$. Then, one can go over all $y_{1:n} \in \mathcal{X}^n$, and use the explicit constructions from the proof of Lemma B.2 regarding the product, sum, and absolute value, yielding rational-valued computable upper and lower approximations of $f$. □

We know that $A_{n\,n\geq 1}$ is computable. From Lemma B.2, the product of two computable functions is computable, thus lower semi-computable, and the ceiling function preserves semi-computability. Therefore, $\delta$ is lower semi-computable. Since $\delta - 2\log(\delta + 3) = \log(2^\delta/(\delta+3)^2)$, then by Lemma B.2, for any positive integer $L$, function $\delta - 2\log(\delta + 3) - L$ is lower semi-computable. It remains to prove that a certain choice of $L$ yields a $p_X^{\otimes *}$-critic. From (82) and (90), there exists $N_0 \in \mathbb{N}$ such that

$$\forall n \geq N_0, \ B_n \geq \lambda(|\mathcal{X}|^n + 2)A_n^2. \tag{103}$$

For any $n \geq N_0$, any $C \geq 2$, any integer $t \in [nB_n, (n+1)B_{n+1})$, and any $x_{1:t} \in \mathcal{X}^t$, we have:

$$\{\delta(x_{1:t}) \geq C\} = \left\{ \left\lceil A_n \left\|\mathbb{P}^{emp}_{\mathcal{X}^n}[x_{1:nB_n}] - p^{\otimes n}\right\|_{TV}\right\rceil \geq C \right\}$$
$$= \left\{ A_n \left\|\mathbb{P}^{emp}_{\mathcal{X}^n}[x_{1:nB_n}] - p^{\otimes n}\right\|_{TV} > C - 1 \right\}.$$

From this, (103), and Lemma G.1, with distribution $p_X^{\otimes n}$, and parameters $b = B_n$, $\varepsilon = (C-1)/A_n, \eta = 2^{-C}$ we obtain,

$$\forall t \geq [N_0 B_{N_0}, \infty) \cap \mathbb{N}, \forall C \in \mathbb{N}_{\geq 2}, \quad p_X^{\otimes t}(\delta(X_{1:t}) \geq C) \leq 2^{-C}.$$

Therefore,

$$\forall t \geq [N_0 B_{N_0}, \infty) \cap \mathbb{N}, \forall C \in \mathbb{N}_{\geq 2}, \quad \mathbb{E}_{p_X^{\otimes t}}[\mathbf{1}_{\delta(X_{1:t})=C} \cdot 2^{\delta(X_{1:t})-2\log(\delta(X_{1:t})+3)}] \leq \frac{1}{(C+3)^2}.$$

This also holds for $C \in \{0, 1\}$. Summing over $C \in \mathbb{N}_{\geq 0}$ gives,

$$\forall t \geq [N_0 B_{N_0}, \infty) \cap \mathbb{N}, \quad \sum_{x_{1:t} \in \mathcal{X}^t} p_X^{\otimes t}(x_{1:t}) \cdot 2^{\delta(x_{1:t})-2\log(\delta(x_{1:t})+3)} \leq 1. \tag{104}$$

In order to extend this to all positive integers $t$, it is sufficient to multiply by $2^{-L}$ for some $L$ large enough. Therefore, there exists $L \in \mathbb{N}$ such that $\delta - 2\log(\delta + 3) - L$ is a lower semi-computable $p_X^{\otimes *}$-critic. This concludes the proof.

# J On the total variation distance

## J.1 Some lemmas

**Lemma J.1.** *Let $\Pi$ and $\Gamma$ be two distributions on a set $\mathcal{W} \times \mathcal{L}$. Then*

$$\|\Pi_W - \Gamma_W\|_{TV} \leq \|\Pi_{W,L} - \Gamma_{W,L}\|_{TV}.$$

**Lemma J.2.** *Let $\Pi$ and $\Gamma$ be two distributions on a set $\mathcal{W} \times \mathcal{L}$. Then when using the same conditional probability kernel $\Pi_{L|W}$, we have*

$$\|\Pi_W \Pi_{L|W} - \Gamma_W \Pi_{L|W}\|_{TV} = \|\Pi_W - \Gamma_W\|_{TV}.$$

**Lemma J.3.** *Let $\Pi$ and $\Gamma$ be two distributions on a set $\mathcal{W}$, and $f : \mathcal{W} \to \mathbb{R}$ be a bounded function. Then,*

$$|\mathbb{E}_\Pi[f] - \mathbb{E}_\Gamma[f]| \leq 2 \max |f| \cdot \|\Pi - \Gamma\|_{TV}.$$

## J.2 Proof of Claim D.5

Let $P$ and $Q$ be any two distributions on the same alphabet. Fix a positive integer $B$. Then, we have, with the convention $\Pi \otimes \Gamma^{\otimes 0} \equiv \Pi$,

$$\|P^B - Q^B\|_{TV} = \|\sum_{k=1}^{B} (P^{\otimes(B-k+1)} \otimes Q^{\otimes(k-1)} - P^{\otimes(B-k)} \otimes Q^{\otimes k})\|_{TV}$$

$$\leq \sum_{k=1}^{B} \|P^{\otimes(B-k+1)} \otimes Q^{\otimes(k-1)} - P^{\otimes(B-k)} \otimes Q^{\otimes k}\|_{TV} \tag{105}$$

$$\leq \sum_{k=1}^{B} \|P^{\otimes(B-k)} \otimes P \otimes Q^{\otimes(k-1)} - P^{\otimes(B-k)} \otimes Q \otimes Q^{\otimes(k-1)}\|_{TV}$$

$$\leq \sum_{k=1}^{B} \|P - Q\|_{TV} = B\|P - Q\|_{TV}, \tag{106}$$

where (105) follows from the triangle inequality for the TVD ; and (106) follows form Lemma J.2, with $W = X_{B-k+1}$, $\Pi_W \equiv P$ and $\Gamma_W \equiv Q$.

# K The birthday paradox

We provide a proof of Claim D.2. We have

$$(p_{[\lfloor 2^{R_1} \rfloor]}^{\mathcal{U}})^{\otimes B} (M^{(1)}, ..., M^{(B)} \text{ 2 by 2 distinct}) = \prod_{k=1}^{B} \frac{\lfloor 2^{R_1} \rfloor - k + 1}{\lfloor 2^{R_1} \rfloor}$$

$$\geq \frac{(\lfloor 2^{R_1} \rfloor - B + 1)^B}{\lfloor 2^{R_1} \rfloor^B}$$

$$\geq \left(1 - \frac{B - 1}{\lfloor 2^{R_1} \rfloor}\right)^B$$

$$\geq 1 - \frac{B(B - 1)}{\lfloor 2^{R_1} \rfloor} \tag{107}$$

$$\geq 1 - \frac{B^2}{\lfloor 2^{R_1} \rfloor},$$

where (107) follows from Bernoulli's inequality, since $R_1 > \log(B)$.

# L  Existence of a universal $p^{\otimes *}$-critic

We provide a proof of Theorem B.4. From Theorem 4.3.1 in [14], there exists a sequence $\{q_n\}_{n \geq 1}$ containing all lower semi-computable semi-measures on $\{0,1\}^*$, and a sequence $\{\pi_n\}_{n \geq 1}$ of (strictly) positive reals, such that the mixture defined by

$$\mathbf{m} := \sum_{n \geq 1} \pi_n q_n \tag{108}$$

is a lower semi-computable semi-measure on $\{0,1\}^*$. For every $n \in \mathbb{N}$, let $\mathbf{m}(\mathcal{X}^n)$ denote

$$\sum_{x_{1:n} \in \mathcal{X}^n} \mathbf{m}(x_{1:n}).$$

From (108), we have $\forall x \in \{0,1\}^*, \mathbf{m}(x) > 0$. Moreover, $\forall x_0 \in \mathcal{X}, \ p(x_0) > 0$, thus $\forall x \in \cup_{n \in \mathbb{N}} \mathcal{X}^n, \ p^{\otimes *}(x) > 0$. Define function $\delta_0$, by

$$\forall n \in \mathbb{N}, \forall x_{1:n} \in \mathcal{X}^n, \ \delta_0(x_{1:n}) := \log\Big(\frac{\mathbf{m}(x_{1:n})}{\mathbf{m}(\mathcal{X}^n)p^{\otimes n}(x_{1:n})}\Big). \tag{109}$$

Fix any lower semi-computable $p^{\otimes *}$-critic $\delta$. Define map $q_\delta : \{0,1\}^* \to \mathbb{R}$ by

$$\forall x \in \cup_{n \in \mathbb{N}} \mathcal{X}^n, \quad q_\delta(x) := \mathbf{m}(\mathcal{X}^{l(x)}) 2^{\delta(x)} p^{\otimes *}(x),$$

and $x \mapsto 0$ elsewhere. From Lemma B.2 (iii), the function which is null outside of $\cup_{n \in \mathbb{N}} \mathcal{X}^n$, and defined by $x \mapsto \mathbf{m}(\mathcal{X}^{l(x)})$ on $\cup_{n \in \mathbb{N}} \mathcal{X}^n$, is lower semi-computable. Moreover, $x \mapsto 2^{\delta(x)}$ and $x \mapsto p^{\otimes *}(x)$ are lower semi-computable by Lemma B.2 (i) and (iii) respectively. Thus, $q_\delta$ is the product three non-negative lower semi-computable functions. Hence, $q_\delta$ is lower semi-computable by Lemma B.2 (i). Moreover, we have

$$\sum_{x \in \{0,1\}^*} q_\delta(x) = \sum_{n \in \mathbb{N}} \mathbf{m}(\mathcal{X}^n) \sum_{x \in \mathcal{X}^n} 2^{\delta(x)} p^{\otimes n}(x)$$

$$\leq \sum_{n \in \mathbb{N}} \mathbf{m}(\mathcal{X}^n) \tag{110}$$

$$\leq 1, \tag{111}$$

where (110) follows from the definition of a $p^{\otimes *}$-critic; and (111) follows from the fact that $\mathbf{m}$ is a semi-measure. Therefore, $q_\delta$ is a lower semi-computable semi-measure. Thus, from (108), we have $\mathbf{m} \geq \pi_{q_\delta} q_\delta$, for some positive real $\pi_{q_\delta}$. In order to derive (14), fix $x \in \cup_{n \in \mathbb{N}} \mathcal{X}^n$, and denote $l(x)$ by $n$. From (108), we have $\mathbf{m}(x) > 0$. Therefore, since $\forall x_0 \in \mathcal{X}, p(x_0) > 0$, we have $q_\delta(x) > 0$. Thus, from (109), we have

$$\delta_0(x) = \log\Big(\frac{\mathbf{m}(x_{1:n})}{\mathbf{m}(\mathcal{X}^n)p^{\otimes n}(x_{1:n})}\Big)$$

$$\geq \log\Big(\frac{\pi_{q_\delta} q(x_{1:n})}{\mathbf{m}(\mathcal{X}^n)p^{\otimes n}(x_{1:n})}\Big)$$

$$= \log(\pi_{q_\delta}) + \delta(x).$$

This is true for any lower semi-computable $p^{\otimes *}$-critic $\delta$, and any $x \in \cup_{n \in \mathbb{N}} \mathcal{X}^n$. Since $\log(\pi_{q_\delta})$ does not depend on $x$, then property (14) holds. This concludes the proof.

# M  Additional semi-computability arguments

We provide a proof of Lemma B.2. If $f$ is lower semi-computable, we denote by $(x,k) \mapsto \varphi_{f,-}(x,k)$ a computable function from $\mathcal{E}$ to $\mathbb{Q}$, monotonically approaching $f$ from below, in the sense of Definition B.1. If $f$ is upper semi-computable, then $\varphi_{f,+}(x,k)$ denotes a function of the form $\varphi_{-f,-}(x,k)$, which monotonically approaches $f$ from above.

## M.1  Assume that $f$ and $g$ are computable

### M.1.1  $f+g$

Function $\varphi_{f,-} + \varphi_{g,-}$ is a computable function from $\mathcal{E} \times \mathbb{N}$ to $\mathbb{Q}$, which monotonically approaches $f+g$ from below. Similarly, $\varphi_{f,+} + \varphi_{g,+}$ constitutes a computable rational upper approximation.

### M.1.2  $|f|$

We construct $\varphi_{|f|,-}$ as follows. Let $x \in \mathcal{E}$ and $k \in \mathbb{N}$. If $\varphi_{f,-}(x,k) \geq 0$, return $|\varphi_{f,-}(x,k)|$. Otherwise, if $\varphi_{f,+}(x,k) \leq 0$, return $|\varphi_{f,+}(x,k)|$. Otherwise, return 0. We define $\varphi_{|f|,+}(x,k)$ as

$$\max\Big(|\varphi_{f,-}(x,k)|, |\varphi_{f,+}(x,k)|\Big).$$

Straightforwardly, this implies that $|f|$ is computable.

### M.1.3  $fg$

Define $\varphi_{fg,-}(x,k)$ as follows. If $\varphi_{f,-}(x,k) \geq 0$ and $\varphi_{g,-}(x,k) \geq 0$, then return $\varphi_{f,-}(x,k)\varphi_{g,-}(x,k)$. Otherwise, if $\varphi_{f,+}(x,k) \leq 0$ and $\varphi_{g,+}(x,k) \leq 0$, then return $\varphi_{f,+}(x,k)\varphi_{g,+}(x,k)$. Otherwise, return

$$-\max\Big(|\varphi_{f,-}(x,k)|, |\varphi_{f,+}(x,k)|\Big)\max\Big(|\varphi_{g,-}(x,k)|, |\varphi_{g,+}(x,k)|\Big).$$

Define $\varphi_{fg,+}$ as $-\varphi_{(-f)g,-}$.

## M.2  Suppose that $f$ is computable and only takes positive values

### M.2.1  $1/f$

Define $\varphi_{1/f,-}(x,k)$ as $1/\varphi_{f,+}(x,k)$. Compute $k_1(x)$, the smallest positive integer $k$ such that $\varphi_{f,-}(x,k) > 0$. For all integers $k \in [1, k_1(x)]$, define $\varphi_{1/f,+}(x,k)$ as $1/\varphi_{f,-}(x,k_1(x))$. For all integers $k \in (k_1(x), \infty)$, define $\varphi_{1/f,+}(x,k)$ as $1/\varphi_{f,-}(x,k)$.

### M.2.2  $f^{1/b}$

Compute $k_1(x)$, the smallest positive integer $k$ such that $\varphi_{f,-}(x,k) > 0$. For all integers $k \in [1, k_1(x))$, define $\varphi_{f^{1/b},-}(x,k) := 0$ and $\varphi_{f^{1/b},+}(x,k) = \lceil \varphi_{f,+}(x,k) \rceil$. Consider an integer $k \in [k_1(x), \infty)$. Compute the greatest integer $m$ such that $(m/2^k)^b \leq \varphi_{f,-}(x,k)$. Then, define $\varphi_{f^{1/b},-}(x,k) := m/2^k$. Therefore, we have

$$\forall k \geq k_1(x), \ \ 0 \leq \varphi_{f,-}(x,k)^{1/b} - \varphi_{f^{1/b},-}(x,k) < \frac{1}{2^k}. \tag{112}$$

From (112), and since the $b$-th root function and $k \mapsto \varphi_{f,-}(x,k)$ are both non-decreasing, we have

$$\forall k \geq k_1(x)+1, \ \ \varphi_{f^{1/b},-}(x,k-1) \leq \varphi_{f,-}(x,k)^{1/b}. \tag{113}$$

Since $\varphi_{f^{1/b},-}(x,k-1)$ can also be written in the form $m'/2^k$, then, from the maximality of the integer $m$ appearing in the construction of $\varphi_{f^{1/b},-}(x,k)$, we have

$$\forall k \geq k_1(x)+1, \ \ \varphi_{f^{1/b},-}(x,k-1) \leq \varphi_{f^{1/b},-}(x,k). \tag{114}$$

This also holds for all integers $k \in [2, k_1(x)+1]$. Properties (112) and (114) imply that $f^{1/b}$ is lower semi-computable. We prove upper semi-computability similarly, using the smallest integer $\tilde{m}$ such that $(\tilde{m}/2^k)^b \geq \varphi_{f,+}(x,k)$, and setting $\varphi_{f^{1/b},+}(x,k) := \tilde{m}/2^k$.

## M.3  Assume that $f$ and $g$ are lower semi-computable

### M.3.1  $f+g$

Function $\varphi_{f,-} + \varphi_{g,-}$ is a computable function from $\mathcal{E} \times \mathbb{N}$ to $\mathbb{Q}$, which monotonically approaches $f+g$ from below.

### M.3.2  $\lceil f \rceil$

Define $\varphi_{\lceil f \rceil,-}$ as $\lceil \varphi_{f,-} \rceil$.

### M.3.3  $2^f$

Fix $x \in \mathcal{E}$ and $k \in \mathbb{N}$. Let $a \in \mathbb{Z}$ and $b \in \mathbb{N}$ such that $\varphi_{f,-}(x,k) = a/b$. Compute the greatest integer $m$ such that $(m/2^k)^b \leq 2^a$. Then, define $\varphi_{2^f,-}(x,k) := m/2^k$. Therefore, we have

$$0 \leq 2^{\varphi_{f,-}(x,k)} - \varphi_{2^f,-}(x,k) < \frac{1}{2^k}. \tag{115}$$

From (115), and since the exponential function and $k \mapsto \varphi_{f,-}(x,k)$ are both non-decreasing, we have

$$\forall k \geq 2, \; \varphi_{2^f,-}(x,k-1) \leq 2^{\varphi_{f,-}(x,k)}. \tag{116}$$

Since $\varphi_{2^f,-}(x,k-1)$ can also be written in the form $m'/2^k$, then, from the maximality of the integer $m$ appearing in the construction of $\varphi_{2^f,-}(x,k)$, we have

$$\forall k \geq 2, \; \varphi_{2^f,-}(x,k-1) \leq \varphi_{2^f,-}(x,k). \tag{117}$$

Properties (115) and (117) imply that $2^f$ is lower semi-computable.

### M.4  Assume that $f$ and $g$ are semi-computable and non-negative

### M.4.1  $fg$

If $\varphi_{f,-}(x,k) \geq 0$ and $\varphi_{g,-}(x,k) \geq 0$, return $\varphi_{f,-}(x,k)\varphi_{g,-}(x,k)$. Otherwise, return $0$.

### M.4.2  $2^f/(3+f)^2$

There exists a real $\varepsilon \in (0,1)$ such that $u \mapsto 2^u/(3+u)^2$ is non-decreasing on $(-\varepsilon, \infty)$. Fix $x \in \mathcal{E}$. Compute $k_1(x)$, the smallest positive integer $k$ such that $\varphi_{f,-}(x,k) > -\varepsilon$. For all integers $k \in [1, k_1(x))$, define $\varphi_{2^f/(3+f)^2,-}(x,k) := 0$. Fix an integer $k \geq k_1(x)$. Let $a \in \mathbb{Z}$ and $b \in \mathbb{N}$ such that $\varphi_{f,-}(x,k) = a/b$. Compute the greatest integer $m$ such that $(m/2^k)^b \leq 2^a/(3+a/b)^{2b}$. Then, define $\varphi_{2^f/(3+f)^2,-}(x,k) := m/2^k$. Therefore, we have

$$\forall k \geq k_1(x), \; 0 \leq \frac{2^{\varphi_{f,-}(x,k)}}{(3+\varphi_{f,-}(x,k))^2} - \varphi_{2^f/(3+f)^2,-}(x,k) < \frac{1}{2^k}. \tag{118}$$

From (118), and since $k \mapsto \varphi_{f,-}(x,k)$ is non-decreasing, and $u \mapsto 2^u/(3+u)^2$ is non-decreasing on $(-\varepsilon, \infty)$, we have

$$\forall k \geq k_1(x)+1, \; \varphi_{2^f/(3+f)^2,-}(x,k-1) \leq \frac{2^{\varphi_{f,-}(x,k)}}{(3+\varphi_{f,-}(x,k))^2}. \tag{119}$$

Since $\varphi_{2^f/(3+f)^2,-}(x,k-1)$ can also be written in the form $m'/2^k$, then, from the maximality of the integer $m$ appearing in the construction of $\varphi_{2^f/(3+f)^2,-}(x,k)$, we have

$$\forall k \geq k_1(x)+1, \; \varphi_{2^f/(3+f)^2,-}(x,k-1) \leq \varphi_{2^f/(3+f)^2,-}(x,k). \tag{120}$$

This is also true for all integers $k \in [2, k_1(x)+1]$. Properties (118) and (120) imply that $2^f/(3+f)^2$ is lower semi-computable.

### M.4.3  $\log(f)$

Assume that $f$ only takes positive values. Fix $x \in \mathcal{E}$. Compute $k_1(x)$, the smallest positive integer $k$ such that $\varphi_{f,-}(x,k) > 0$. Fix an integer $k \geq k_1(x)$. Compute the largest integer $m$ such that $2^m \leq \varphi_{f,-}(x,k)^{2^k}$. Then, define $\varphi_{\log(f),-}(x,k) := m/2^k$. For all integers $k \in [1, k_1(x))$, define $\varphi_{\log(f),-}(x,k)$ as $\varphi_{\log(f),-}(x,k_1(x))$. Therefore, we have

$$\forall k \geq k_1(x), \; 0 \leq \log(\varphi_{f,-}(x,k)) - \varphi_{\log(f),-}(x,k) < \frac{1}{2^k}. \tag{121}$$

From (121), and since the logarithm and $k \mapsto \varphi_{f,-}(x,k)$ are both non-decreasing, we have

$$\forall k \geq k_1(x)+1, \; \varphi_{\log(f),-}(x,k-1) \leq \log(\varphi_{f,-}(x,k)). \tag{122}$$

Since $\varphi_{\log(f),-}(x,k-1)$ can also be written in the form $m'/2^k$, then, from the maximality of the integer $m$ appearing in the construction of $\varphi_{\log(f),-}(x,k)$, we have

$$\forall k \geq k_1(x)+1, \; \varphi_{\log(f),-}(x,k-1) \leq \varphi_{\log(f),-}(x,k). \tag{123}$$

This also holds for all integers $k \in [2, k_1(x)+1]$. Properties (121) and (123) imply that $\log(f)$ is lower semi-computable.

## M.5 Functions of finite binary strings

Let $\mathcal{X}$ be a finite computable subset of $\{0, 1\}^*$, and $f$ be a lower semi-computable function from $\{0, 1\}^*$ into $\mathbb{R}$.

**Lemma M.1.** $\cup_{n \in \mathbb{N}} \mathcal{X}^n$ *is a computable set.*

*Proof.* By Definition B.1, it is sufficient to construct a computable function $\tau$ from $\{0, 1\}^*$ to $\{0, 1\}$, which returns 1 if its input is in $\cup_{n \in \mathbb{N}} \mathcal{X}^n$, and 0 otherwise. Since $\mathcal{X}$ is computable, there exists a computable function $\tau_0$ from $\{0, 1\}^*$ to $\{0, 1\}$, which returns 1 if its input is in $\mathcal{X}$, and 0 otherwise. Fix $x \in \{0, 1\}^*$. Define $\tau(x)$ as follows. Enumerate all partitions of $x$ into consecutive sub-strings. For each, call $\tau_0$ on every sub-string. If for some partition, the output of $\tau_0$ is 1 for every sub-string, then return 1. Otherwise, return 0. $\qquad\square$

Hereafter, we use the notation $\tau$ defined in the above proof.

### M.5.1 Partial sums

Consider the function $\tilde{f} : \{0, 1\}^* \to \mathbb{R}$ which is null outside of $\cup_{n \in \mathbb{N}} \mathcal{X}^n$, and is defined by

$$\forall x \in \cup_{n \in \mathbb{N}} \mathcal{X}^n, \quad \tilde{f}(x) = \sum_{y \in \mathcal{X}^{l(x)}} f(y).$$

Fix $x \in \{0, 1\}^*$ and $k \in \mathbb{N}$. Define $\varphi_{\tilde{f}, -}(x, k)$ as follows. Compute $\tau(x)$. If it is null, return 0. Otherwise: compute $l(x)$, and for each $y$ in $\mathcal{X}^{l(x)}$, compute $\varphi_{f, -}(y, k)$, then return

$$\sum_{y \in \mathcal{X}^{l(x)}} \varphi_{f, -}(y, k).$$

Fix some $x \in \cup_{n \in \mathbb{N}} \mathcal{X}^n$. The (finite) set of indices of the above sum does not depend on $k$. Therefore, since for each $y \in \{0, 1\}^*$ we have $\varphi_{f, -}(y, k) \underset{k \to \infty}{\to} f(y)$, we get

$$\forall x \in \cup_{n \in \mathbb{N}} \mathcal{X}^n, \quad \varphi_{\tilde{f}, -}(x, k) \underset{k \to \infty}{\longrightarrow} \tilde{f}(x). \tag{124}$$

Similarly, since for any $y \in \{0, 1\}^*$ and any $k \geq 1$, we have $\varphi_{f, -}(y, k) \leq \varphi_{f, -}(y, k + 1)$, then we have

$$\forall x \in \cup_{n \in \mathbb{N}} \mathcal{X}^n, \forall k \in \mathbb{N}, \quad \varphi_{\tilde{f}, -}(x, k) \leq \varphi_{\tilde{f}, -}(x, k + 1). \tag{125}$$

Properties (124) and (125) also hold for finite strings outside of $\cup_{n \in \mathbb{N}} \mathcal{X}^n$. Thus, $\tilde{f}$ is lower semi-computable.

### M.5.2 Product distribution

Let $p$ be a lower semi-computable probability measure on $\mathcal{X}$. Fix $x \in \{0, 1\}^*$ and $k \in \mathbb{N}$. Define $\varphi_{p^{\otimes *}, -}(x, k)$ as follows. Compute $\tau(x)$. If it is null, return 0. Otherwise, proceed as follows. Compute $l(x)$. We write $x$ as $x_{1:l(x)}$, with $x_t \in \mathcal{X}$ for any integer $t$ in $[1, l(x)]$. Compute and return

$$\prod_{t=1}^{l(x)} \varphi_{p, -}(x_t, k).$$

Fix some $x \in \cup_{n \in \mathbb{N}} \mathcal{X}^n$. The (finite) set of indices of the above product does not depend on $k$. Therefore, since for each $y \in \mathcal{X}$, we have $\varphi_{p, -}(y, k) \underset{k \to \infty}{\to} p(y)$, we get

$$\forall x \in \cup_{n \in \mathbb{N}} \mathcal{X}^n, \quad \varphi_{p^{\otimes *}, -}(x, k) \underset{k \to \infty}{\longrightarrow} p^{\otimes *}(x). \tag{126}$$

Similarly, since for any $y \in \mathcal{X}$, and any $k \geq 1$, we have $\varphi_{p, -}(y, k) \leq \varphi_{p, -}(y, k + 1)$, and since $p$ is non-negative, then we have

$$\forall x \in \cup_{n \in \mathbb{N}} \mathcal{X}^n, \forall k \in \mathbb{N}, \quad \varphi_{p^{\otimes *}, -}(x, k) \leq \varphi_{p^{\otimes *}, -}(x, k + 1). \tag{127}$$

Properties (126) and (127) also hold for finite strings outside of $\cup_{n \in \mathbb{N}} \mathcal{X}^n$. Thus, $p^{\otimes *}$ is lower semi-computable.