

# The Bateson One-Way Function: Formal Definition, Cryptographic Hardness, and Information-Theoretic Irreversibility

Kevin Fathi

fathikevin@protonmail.com  
ORCID: 0009-0001-5546-1475

August 2025

## Abstract

We present a rigorous formalization of the Bateson One-Way Function, a novel framework for cryptographic irreversibility rooted in symbolic grammar theory rather than number theory. We define the function based on a formal grammar comprising messages, interpretive frames, and a canonicalization process. Irreversibility arises from "frame suppression," where the context of interpretation is discarded. We establish the function's potential for cryptographic hardness through two distinct arguments. First, we define the Semantic Inference Problem (SIP)—the difficulty of recovering a hidden frame—and prove its average-case hardness via a reduction from Planted 3-SAT. This demonstrates that the Bateson framework can encode hard inference problems. Second, we characterize the conditions required for a Bateson instantiation to be a secure One-Way Function (OWF) using the established equivalence between OWFs and the hardness of time-bounded Kolmogorov complexity (pKt). We introduce a taxonomy distinguishing between ambiguity-based (compressing) and complexity-based (expanding) Bateson functions, and quantify information loss using Symbolic Degeneracy ( $\Delta_G$ ). This work solidifies the Bateson function's theoretical foundation as a candidate for non-algebraic, post-quantum cryptography.

## 1 Introduction

One-way functions (OWFs) are a cornerstone of cryptography. Traditional designs rely on the assumed hardness of number-theoretic problems. The Bateson One-Way Function proposes a novel approach: irreversibility rooted in the suppression of semantic context. Inspired by Gregory Bateson's concepts and formalized via symbolic grammar theory, this function achieves one-wayness through the suppression of the interpretive context (the "frame").

This paper provides a rigorous definition of the function and explores two independent theoretical foundations for its security. We analyze the hardness of inferring the suppressed context (the Semantic Inference Problem) and characterize the structural properties required for cryptographic one-wayness based on modern meta-complexity results.

## 2 Preliminaries

We denote the security parameter by  $n$ . PPT stands for Probabilistic Polynomial Time.

**Definition 1** (One-Way Function (OWF)). *A function  $f : \{0, 1\}^* \rightarrow \{0, 1\}^*$  is a strong one-way function if it is easy to compute (PPT) and hard to invert: For every non-uniform PPT adversary  $\mathcal{A}$ , there exists a negligible function  $\mu(\cdot)$  such that [5]:*

$$\Pr[x \leftarrow \{0, 1\}^n; y \leftarrow f(x) : f(\mathcal{A}(1^n, y)) = y] \leq \mu(n)$$

**Definition 2** (Time-Bounded Kolmogorov Complexity (Kt)). *The  $t$ -bounded Kolmogorov complexity of a string  $y$ , denoted  $K^t(y)$ , is the length of the shortest program  $p$  that outputs  $y$  on a fixed universal Turing machine  $U$  within  $t(|y|)$  steps [6].*

It is established that the existence of OWFs is equivalent to the average-case hardness of computing  $K^t(y)$  for efficiently samplable distributions (the pKt problem) [8, 9].

### 3 Definition of the Bateson One-Way Function

The Bateson function is defined over a symbolic grammar that specifies how messages are interpreted under different contexts.

#### 3.1 The Bateson Grammar and Function

**Definition 3.** *A Recursive Bateson Grammar is a tuple  $G = (M, F, \Theta, R, CNF_G)$ , where  $M$  is the message space,  $F$  is the set of interpretive Frames,  $\Theta$  is the output space,  $R$  defines the interpretation Rules  $M^+ \times F \rightarrow \Theta^*$ , and  $CNF_G$  is an efficiently computable Canonicalization Function mapping raw outputs to a unique normal form in  $\Theta^+$ .*

**Definition 4.** *Given a grammar  $G$ , the Bateson function  $f_B^+ : M^+ \times F \rightarrow \Theta^+$  is defined as:*

$$f_B^+(m, f) := CNF_G(R(m, f))$$

We assume  $f_B^+$  is efficiently computable.

The crucial characteristic is **Frame Suppression**. The output  $y$  is produced, but the frame  $f$  is discarded.

#### 3.2 Quantifying Ambiguity: Symbolic Degeneracy

**Definition 5.** *The Symbolic Degeneracy  $\Delta_G(y)$  of an output  $y \in \Theta^+$  is the logarithmic size of its preimage set under the grammar  $G$ .*

$$\Delta_G(y) := \log_2 |\{(m, f) \in M^+ \times F : f_B^+(m, f) = y\}|$$

$\Delta_G(y)$  quantifies the information, in bits, lost during the forward computation.

#### 3.3 A Taxonomy of Bateson Functions

The cryptographic properties of  $f_B^+$  depend crucially on the characteristics of the grammar  $G$ . We distinguish two primary classes:

1. **Ambiguity-Based (Compressing):** Grammars where the output space is significantly smaller than the input space. These functions exhibit high Symbolic Degeneracy. Security relies on the hardness of finding a preimage despite many collisions, similar to cryptographic hash functions.
2. **Complexity-Based (Expanding):** Grammars where the function is length-expanding and nearly injective (low or bounded Symbolic Degeneracy). Security relies on the computational complexity (incompressibility) of the output, ensuring the input cannot be efficiently recovered.

### 4 Hardness of Context Inference (SIP)

We first investigate the difficulty of recovering the interpretive frame. This is formalized as the Semantic Inference Problem (SIP).

## 4.1 The Semantic Inference Problem

**Definition 6** (Semantic Inference Problem (SIP)). ***Input:** A grammar  $G$ , a distribution  $\mathcal{D}$  over  $M^+$ , and samples  $\{(m_i, y_i)\}$  where  $m_i \sim \mathcal{D}$  and  $y_i = f_B^+(m_i, f^*)$  for an unknown, fixed frame  $f^* \in F$ .*

***Output:** Identify the frame  $f^*$ .*

SIP is an average-case learning problem. Its hardness implies the frame can function as a cryptographic key.

## 4.2 Theorem: SIP is Hard on Average

We prove that there exists a grammar construction for which SIP is hard, based on the assumed hardness of the Planted 3-SAT problem [7].

**Theorem 1.** *If the Planted 3-SAT problem is hard on average, then SIP is hard on average for a specific, constructible grammar  $G_{SAT}$ .*

*Proof.* We construct a grammar  $G_{SAT}$  that embeds a 3-SAT instance with  $n$  variables and  $m$  clauses.

**1. Grammar Construction ( $G_{SAT}$ ):** Frames ( $F = \{0, 1\}^n$ ) represent truth assignments  $A$ . Messages ( $M^+$ ) represent clauses  $C_j$ . Outputs ( $\Theta = \{1, 0\}$ ). Rules ( $R$ ):  $f_B^+(m_j, f_A)$  evaluates clause  $C_j$  under assignment  $A$ .

**2. The Reduction:** Given a Planted 3-SAT instance with hidden assignment  $A^*$ . The hidden frame is  $f^* = f_{A^*}$ . Since  $A^*$  satisfies all clauses, the samples are  $\{(m_1, 1), \dots, (m_m, 1)\}$ . The SIP task is to find  $f^*$ .

**3. Hardness:** An algorithm that solves this SIP instance recovers  $f^*$ , which corresponds to  $A^*$ . Therefore, if Planted 3-SAT is hard, this construction of SIP is hard.  $\square$

## 4.3 Implications and Limitations (SIP vs. OWF)

Theorem 4.2 demonstrates that the Bateson framework is expressive enough to encode hard average-case problems. However, it is crucial to note that this does not prove  $G_{SAT}$  is a OWF.

$G_{SAT}$  is an Ambiguity-Based (Compressing) construction; the output is only a single bit. A OWF requires that it is hard to find \*any\* preimage  $(m', f')$  given an output  $y$ . In  $G_{SAT}$ , inverting the function means finding \*any\* clause and assignment that satisfy it (e.g., given output 1). This is computationally trivial.

The hardness established here concerns recovering the specific, hidden assignment (the frame) that satisfies all constraints simultaneously. This is relevant for constructing keyed primitives, but establishing a Bateson OWF requires different properties.

## 5 One-Wayness via Meta-Complexity

We now address the standard definition of a One-Way Function, focusing on the Complexity-Based (Expanding) class, utilizing the equivalence between OWFs and the hardness of pKt.

### 5.1 Characterization of Complexity-Based OWFs

We define the properties required for a Bateson function to be a secure OWF based on computational incompressibility.

**Definition 7** (Computational Compression-Resistance). *A polynomial-time, length-expanding function  $f$  is computationally compression-resistant if for a random input  $x$  of length  $N$ , the output  $y = f(x)$  satisfies, with overwhelming probability:*

- 1. High Output Complexity:** *The output is computationally incompressible. For any polynomial  $t(N)$ ,  $K^t(y) \geq |y| - o(N)$ .*

2. **Bounded Symbolic Degeneracy:**  $\Delta_G(y)$  is bounded (e.g., by a polynomial in  $N$ ).

**Theorem 2.** Assume the average-case hardness of  $pKt$  (i.e., OWFs exist). If an instantiation of the Bateson function  $f_B^+$  is computationally compression-resistant, then  $f_B^+$  is a secure one-way function.

*Proof.* The argument relies on showing that an efficient inverter implies an efficient compressor, contradicting the properties of the function under the  $pKt$  assumption. Let  $x = (m, f)$  be the input and  $y = f_B^+(x)$ .

1. Assume for contradiction that an efficient (PPT) inverter  $\mathcal{I}$  exists. Given  $y$ ,  $\mathcal{I}(y)$  finds a preimage  $x'$  in polynomial time  $t$ .
2. Since the function is length-expanding, the preimage  $x'$  is shorter than the output  $y$  ( $|x'| < |y|$ ).
3. This short preimage  $x'$  can serve as a compact program to generate  $y$  (by computing  $f_B^+(x')$  efficiently). This implies the time-bounded complexity  $K^t(y)$  is low:  $K^t(y) \leq |x'| + O(1)$ .
4. The existence of an efficient inverter  $\mathcal{I}$  therefore implies the existence of an efficient compressor for the function's output distribution.
5. This contradicts the requirement that the output has high time-bounded complexity (Definition 5.1, property 1).
6. Therefore, no efficient inverter  $\mathcal{I}$  can exist, and  $f_B^+$  must be a one-way function.

The condition of bounded degeneracy ensures the function is sufficiently injective, focusing the hardness on the computational irreducibility of the mapping rather than excessive collisions.  $\square$

## 5.2 Implications

Theorem 5.2 provides a roadmap for constructing provably secure Bateson OWFs: design grammars that are provably length-expanding, have bounded degeneracy, and produce computationally incompressible outputs (e.g., using Pseudorandom Functions [3]).

## 6 Empirical Illustration (Toy Example)

We use a simple pedagogical grammar  $G_{Canon}$  to illustrate the Ambiguity-Based mechanism.  $M = \{a, b, c, meta\}$ ,  $F = \{f_1, f_2\}$ . Rules are defined such that ambiguity exists (e.g.,  $f_1(b) = \theta_2$  and  $f_2(a) = \theta_2$ ).

### 6.1 Simulation and Entropy Analysis

We simulate evaluations with uniform frames  $F$  and message distribution  $P(M) = [0.3, 0.3, 0.3, 0.1]$ .

Input Entropy:  $H(M) \approx 1.8955$  bits.  $H(F) = 1$  bit. Total input entropy  $H(M, F) = 2.8955$  bits.

Output Entropy (Simulated):  $H(Y) \approx 1.8955$  bits.

Information Loss:  $H(M, F) - H(Y) \approx 1.0$  bit.

This confirms that the 1 bit of information corresponding to the choice of frame is suppressed, matching the Symbolic Degeneracy  $\Delta_G(y) = 1$ . This illustrates information loss in a compressing function, though it provides no evidence of computational hardness.

## 7 Conclusion

We have presented a unified and rigorous framework for the Bateson One-Way Function. We introduced a taxonomy to classify Bateson functions as Ambiguity-Based (compressing) or Complexity-Based (expanding) and established two distinct theoretical foundations for security within the framework.

We proved that the Semantic Inference Problem (SIP) can be hard (via reduction from Planted 3-SAT), demonstrating the framework’s capability to model hard inference problems. Furthermore, we characterized the conditions (Computational Compression-Resistance) required for a Complexity-Based instantiation to be a secure OWF, relying on the equivalence between OWFs and the hardness of pKt. This work provides the theoretical underpinning for the Bateson function as a viable approach to non-algebraic, post-quantum cryptography.

## References

- [1] [1] S. Arora and B. Barak, Computational Complexity: A Modern Approach, Cambridge University Press, 2009.
- [2] [2] K. Fathi, Bridging Shannon and Turing: A Formal Entropy-Complexity Correspondence over Symbolic Structures, 2025.
- [3] [5] O. Goldreich, Foundations of Cryptography, Cambridge University Press, 2001.
- [4] [6] M. Li and P. Vitányi, An Introduction to Kolmogorov Complexity and Its Applications, Springer, 3rd ed., 2008.
- [5] [7] D. Gamarnik and M. Sudan, Limits of local algorithms for the planted 3-SAT problem, Computational Complexity, 2017. (Representative reference).
- [6] [8] S. Hirahara, et al., One-Way Functions and the Hardness of pKt. ECCC TR24-136 (2024).
- [7] [9] Y. Liu and R. Pass, On One-way Functions and Kolmogorov Complexity. In: FOCS (2020).